Vol. 4, Part 1 Jan.-March 1961

LANGUAGE and SPEECH

edited by **D. B. FRY** with the assistance of
Frieda Goldman-Eisler P. Denes A. C. Gimson

CONTENTS

ROBERT DRAPER LTD.

Kerbihan House, 85 Udney Park Road, Teddington, Middlesex, England

Price £4 (or $11.50) per year (4 issues)

Single copies £1. 5s. (or $4)

Language and Speech

Language and Speech will be published quarterly (March, June, September and December), one volume per annum. The annual subscription is £4 (or $11.50).

Notes for Contributors

Papers are published in English only. Authors submitting material for consideration are asked to comply with the following requirements :

Typescript, which should be the original and not a carbon copy, should be double-spaced, with wide margins. The title of the paper, the author's name and initials, and the name of the institution or organization with which he is connected should be given.

Summary. A short summary should be supplied on a separate sheet and this will be printed at the beginning of the paper.

Sub-headings. Appropriate sub-headings should be inserted in the body of the paper.

References. A complete list of sources referred to in the paper should be provided on separate sheets, and in the following form :

SCRIPTURE, E. W. (1904). The Elements of Experimental Phonetics (New York).

FLETCHER, H., WEGEL, R. L. (1922). The frequency-sensitivity of normal ears. *Phys. Rev.,* 19, 553.

(Note that the titles of articles are required as well as volume and page numbers. In the text of the paper, references should be shown by giving in brackets the surname only of the author and the year of publication thus : (Scripture, 1904).

Phonetic Symbols should be restricted to those used by the International Phonetic Association and their occurrence in the text should be marked by the insertion of oblique strokes thus : /p, t, k/.

Figures should be large line drawings in Indian ink. All figure legends should be typed together on a separate sheet and numbered. The approximate position of figures and tables should be indicated in the text.

Reprints. Fifty reprints will be sent free to contributors and additional copies will be supplied at a reasonable charge.

Contributions should be addressed to the Editor :

D. B. Fry,
Department of Phonetics,
University College,
Gower Street,
London, W.C.1.

Subscriptions should be sent to the Publishers :

Robert Draper Ltd.,
Kerbihan House,
85 Udney Park Road,
Teddington,
Middlesex,
England.

r and
1.50).

ration

ouble-
nitials,
hould

vill be

paper.
ed on

(New

l ears.

nbers.
ackets
pture,

onetic
ion of

typed
igures

copies

o the

ERRATUM

January - March, 1961 issue. We regret the omission of the concluding line of Dr. Eisler's paper on 'Comparative Study of Two Hesitation Phenomena'.

The words printed below should be inserted on p. 26 above the heading 'References'. done, Oct. 8 1964.

SOME CHARACTERISTICS OF WORD CLASSIFICATION

D. J. Bruce
University of Reading

An exploratory study is reported which investigates the effect of given structure on word classification. Subjects have to complete, by selection from a number of alternative items, a word list whose initial entries are systematically varied in relative position from subject to subject. The alternative items fall into three reference categories, Vegetables, Birds and Mammals, but only two are represented by the given entries. The hypothesis that the positional relation of the given words will influence completion strategy is confirmed, and there is some indication of the effect of increasing the initial representation of one of the given topics. The relation between use of the given topic and separation from the given item in subjects' completions is described, and attention is drawn to an underlying consistency in the grouping shown by many of the classifications.

Introduction

The acquisition of language skill by an individual is largely a function of two activities :—

(a) Learning the rules for the serial structure of expression of that language, and
(b) "Cataloguing" or classifying the linguistic units employed.

We shall be concerned here only with the latter activity. Many schemes of classification are used. Some relate to structure, some to length, some to sound. But where higher-order linguistic units — such as words — are concerned, the operationally favoured principle of classification is that of *reference function*. By this term is implied the grouping of units according to their semantic availability for a given reference task. The use of the principle is most obviously demonstrated in the case of word units, and the illustrations following will be confined to this level.

A good formal example of classification by reference function is provided by Roget's Thesaurus. Such a semantic grouping — of equivalence and antithesis within a super-, co-, and sub-ordinate framework — can be made by every language user, though generally to a more limited extent. In seeking referential expression we are all able to draw, more or less readily, on a group of alternative words related by connotation, without the intrusion of irrelevant items. But the associative connections formed between words in individual language experience present a more dynamic picture than any one such formal statement can give. The Thesaurus may be richer in items, but it only begins to suggest the diversity of verbal inter-connections which governs actual language behaviour. In practice, classification by reference function entails much more

than the division of the language into words with a common meaning. Attempts to assess the relations employed by individuals in linking words together (via association tests — see, for example, Karwoski and Berthold, 1945 ; Karwoski and Schachter, 1948) have indicated connections such as " Completion ", " Egocentrism ", " Word derivatives ", and " Predication " which, although they have no place in the Thesaurus, undoubtedly play an important part in cementing ties between words for the language user. " Similarity ", " Contrast ", " Subordinates ", " Co-ordinates ", " Super-ordinates ", " Part-whole ", " Completion ", " Egocentrism ", " Word derivatives ", " Predication " — all these associative directions, and probably more, are actually used in grouping by reference function.

Although so many avenues are open, the results of verbal association tests also suggest that, in certain respects, the individual comes to use an habitually preferred route for word linkage. In the first place, habits of verbal association are formed whereby the linkage between particular words is highly probable. This is not a matter of idiosyncratic experience alone ; habitual responses are likely to be common to defined *groups* of users of a common language. There is evidence of consistency in words related right the way up from the individual re-test, through the astonishing similarity of response amongst members of the same family that Jung (1917) noted, to the communality of association shown by members of a specific professional group (Foley and Macmillan, 1943), and ultimately to the majority of speakers of a common language. Thus in Kent and Rosanoff's (1910) test of 1,000 men and women, 650 gave as their association to the word LAMP the response LIGHT.

More germane to our present interest is a second suggestion of uniformity coming from association test results. This relates character of response to the implicit semantic ties between a number of test words. Howes and Osgood (1954) asked their subjects to make associations to stimulus words, each of which was presented several times, preceded by different word groups. For example, the word DARK would, on one occasion, be prefaced by the cluster DEVIL-FEARFUL-SINISTER. On others, the preceding constellation would take the form of DEVIL-FEARFUL-BASIC, or DEVIL-EAT-BASIC, or 429-124-413. As can be seen, the experimenters' intention was to provide stimulus word groups of varying contextual strength but having the same termination — in this instance the item DARK. It could be argued that, when DARK was placed after DEVIL-FEARFUL-SINISTER, the word formed part of an associative cluster with a unitary reference character, but in the company of digits it shared no semantic ties. The other versions could be regarded as of intermediate strength. Howes and Osgood had previously obtained associations to the group DEVIL-FEARFUL-SINISTER, and to the word DARK alone. Responses to DARK were only occasionally common to those given to the three word group, whose own associates comprised BAD, EVIL, FEAR, FRIGHT, GHOST, etc. But when DARK was placed after these three words and its immediate association requested, 34% of the responses obtained corresponded to those previously given to the prefacing cluster alone. Moreover, the percentage correspondence decreased in step with the assumed strength of contextual influence : 34% when prefaced by DEVIL-FEARFUL-SINISTER, 22% when prefaced by DEVIL-FEARFUL-BASIC, 10% with DEVIL-EAT-BASIC, and

5%, equivalent to that for the word DARK by itself, when the preceding group consisted entirely of digits or nonsense syllables.

Such results emphasise the pervasive influence of semantic grouping on language response, and a further demonstration comes from the work of the present writer (Bruce, 1956). In this case, response took the form of written identifications of test words heard in the presence of decreasing amounts of noise. Four matched lists of monosyllabic words, each fifty items long, were employed as material. Two of the lists were a random selection possessing no associative constraint ; in the other two the items were grouped by community of reference, one list consisting of food words, another of parts of the body. Subjects heard each list six times at improving signal to noise ratios. A significantly superior performance (average of 25% better) was shown on the contextual lists for the four intermediate testing levels used (-1 db, +3 db, +7 db and +12 db). It should be noted that no prior information was given to subjects about the character of the lists they were to hear. Where a constraint existed, and what form the constraint took, was discovered by the subject for himself in the course of testing. In addition to the results reported elsewhere, an interesting feature emerged concerning the point at which a subject first suspected the existence of a constraint. At some stage in the testing subjects would, after completion of a trial, make a remark which indicated their conclusion that all the list items referred to the same topic. On several occasions this was done when the number of previously written responses actually within the (generally correct) context defined was small. Or the decision was referred to a point in the preceding trial where only a few responses had yet been made. Moreover, the contextually appropriate responses that had been made did not necessarily occur in immediate succession. Two examples may be given :—

Subject 8—Context A (Food)—Trial 2: "I did not recognize any of the words as having been heard on Trial 1. I soon grasped the food association—in fact *after about word number five*." The first five of this subject's fifty responses on Trial 2 were, in fact, food words.

Subject 6—Context B (Parts of the Body)—Trial 2: "Are they all parts of the body ? *After six*, I attempted to make them so." In Trial 1 this subject had given nine body words, generally distributed amongst the forty-six responses made. The first six of the fifty responses to Trial 2 contained four body words, again distributed, two of which had also been given in Trial 1.

The most obvious explanation of the feature would be that subjects, having previously encountered one of the contextual lists, might well anticipate another, and hence be prepared to make a decision of constraint on a much-reduced basis of immediate evidence. Against this is the consideration that they might equally well expect another random list, and, in any case, the occurrence of the feature was not confined to experience of the second contextual list. In the examples quoted the subject is describing his impressions on a first experience of a constrained list. Another explanation would be that, although the surprising readiness to make a comprehensive classification did not necessarily result from a specific prior contextual experience, the occurrence of a constrained list is the sort of thing a subject might reasonably expect in a psychological experiment of this kind. No support for this argument could be

found when subjects were questioned at the end of the experiment. A third possibility is that during the trial subjects might have had in mind interpretations of test words which differed from those actually written down in that they were within the operative context. Such an experience might have been combined with the sight of the few contextually appropriate written responses to give rise to the conclusion that a single topic was represented.

Whatever the explanation, and there are other possibilities, the fact remains that the statement by these subjects of their suspicion of an all-embracing topic and the nature of the responses they had actually made did not tally. Frequently the decision that the fifty test items belonged to a common class was made on the immediate basis of from four to ten such written responses. The observation gives rise to the question : " How many items constituting a semantically definable sub-set must there be before the percipient attributes that definition to the whole set ? "

An experimental approach

An experiment is in progress which may give the answer for one set of conditions. The method chosen involves the subjects' own classification of a potential word list by the selection of items from a number of possible alternatives. The alternatives are the same for all subjects, but the classificatory task begins at a different starting point, as defined by given structure, in each case. Thus in the first test to be described a list of twenty possible words is indicated by the initial letters of those words (randomly selected). For every subject the first word of the list has already been filled in — by the same item, TURNIP. One other word is also given, its position being systematically varied from space two to space twenty. This is the name of a bird. The alternatives from which subjects are to select are arranged in alphabetical order and consist of the names of Vegetables, Birds and Mammals. It is possible to give an item from each of these three classes for every initial letter without duplication.

Conditions are the same in the second test except that a second bird name is given, consistently the last item in the list, and the same word — PETREL. Further tests will be based on a given structure of one vegetable, three birds, one vegetable, four birds, and so on. The intention is to see whether a quantitatively definable point emerges where the classificatory balance tips in favour of one topic, and subjects complete the list only with bird names. Collateral tests are planned favouring the other topics, using different topics, having a different number of alternative topics, varying the position of consistently given items, and involving a greater number of list items.

In general, the hypotheses being tested are :—

(a) That the number of given items constituting a semantically definable sub-set will affect classificatory decisions.

(b) That the positional relation of given items in different semantic categories will affect classificatory decisions.

The possibility also exists that a *particular* number of given items referring to a common topic will produce a recognisable change in classificatory decision.

The present experimental report is on an exploratory study, and is concerned principally with hypothesis (b).

Subjects and instructions

Nineteen subjects took part in the first test, and eighteen in the second. They were mostly graduate or undergraduate members of the Psychology Department. Each subject was asked to complete the list " in the way he thought most appropriate " by selecting items from the sheet of alternatives. It was emphasised that the words supplied must start with the initial letters given, and subjects were also told that they could ask about the meaning of any alternatives which were unfamiliar. The request was made that, when the list had been completed, a description should be written of how the subject had approached the task. No time limit was imposed, and subjects were discouraged from thinking that there was any pre-determined way of completing the list. Queries relating to the permissibility of using the same word more than once were answered by " it's entirely up to you ".

Results

Test 1. Given—One Vegetable, One Bird

The results of the first test gave some encouragement to the hypothesis that " completion strategy " would vary with given structure. Table 1 lists the responses, by topic, for the nineteen subjects in this test.

Allowing for the vagaries attendant upon a small sample of this kind, some element of patterning appears. With only one exception (Subject 17), subjects towards the ends of the response matrix, i.e. those for whom the given items were separated by a minimum or maximum of spaces, have completed by :—

(a) Sequential associations, e.g. making up a story, one word suggesting the next, etc. Or,

(b) Topic alternations. Or,

(c) Complicated number schemes.

Completion by runs of responses restricted to the semantic class of the given items is, in the main, limited to subjects at the centre of the matrix. This point will be returned to later.

The general distribution of responses according to topic is not without interest. In all, three hundred and forty-two responses are involved. The percentage division of this total is :—

Vegetable : 43·9%, Bird : 37·4%, Mammal : 18·7%, showing clearly the influence of the given items.

If we call all those responses occurring before the second given item " Pre-change

TABLE 1

Subjects

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
	1	**V**	**V**	**V**	**V**	**V**	**V**	**V**	**V**	**V**	**V**	**V**	**V**	**V**	**V**	**V**	**V**	**V**	**V**	**V**
	2	**B**	M	V	B	V	V	V	V	V	B	V	V	V	V	V	V	V	V	B
	3	V	**B**	M	M	V	V	V	V	M	M	V	V	V	V	V	V	V	V	V
	4	M	V	**B**	V	V	V	V	V	V	B	V	V	V	V	V	V	V	V	B
	5	B	M	B	**B**	V	V	V	V	M	M	V	B	V	V	B	B	V	B	V
	6	B	B	M	M	**B**	V	V	V	B	M	V	M	V	V	B	B	V	B	B
	7	V	V	M	V	B	**B**	V	V	V	V	V	B	V	V	V	B	V	B	V
	8	M	M	V	B	B	B	**B**	V	M	B	V	M	V	V	V	B	V	B	B
	9	B	B	V	M	B	B	V	**B**	B	M	V	B	V	V	V	V	V	M	V
Items	10	B	V	V	V	B	B	V	B	**B**	V	V	B	B	V	B	V	V	M	B
	11	B	M	V	B	M	B	V	B	B	**B**	V	V	B	B	B	V	V	M	V
	12	M	B	V	M	M	B	V	B	B	M	**B**	V	B	B	V	V	V	M	B
	13	B	V	B	V	M	B	V	B	M	V	B	**B**	B	B	V	B	V	V	V
	14	M	M	M	B	M	M	V	B	V	B	B	B	**B**	B	M	B	V	V	B
	15	M	B	M	M	M	M	B	B	M	M	B	B	B	**B**	B	V	V	V	V
	16	V	V	M	V	B	M	V	B	M	V	B	B	B	B	**B**	V	V	V	B
	17	M	M	M	B	B	M	V	B	B	B	B	B	B	B	V	**B**	V	B	V
	18	V	B	M	M	B	M	V	B	M	M	B	B	B	B	V	B	**B**	B	B
	19	V	V	M	V	B	M	V	B	M	V	B	B	B	B	M	V	B	**B**	V
	20	B	M	V	B	B	M	M	B	V	B	B	B	B	B	B	B	B	B	**B**

V=Vegetable; B=Bird; M=Mammal. Given items in **bold type**

Subjects' completions in Test 1

responses ", and those occurring after " Post-change responses ", the respective percentage distribution of topics is as follows :—

Pre-change : Vegetable 64·3%, Bird 25·7%, Mammal 9·9%.

Post-change : Bird 49·1%, Mammal 27·5%, Vegetable 23·4%.

Two inferences may be drawn from these values :—

(a) The first given item (Vegetable) has had a greater constraining effect on the pre-change responses than the second given item (Bird) has had on the post-change responses.

(b) But the effect of the second given item (Bird) has been felt on *the pre-change responses.* Hence the difference in distribution of the two " secondary " topics in the pre-change (Bird and Mammal) and post-change (Vegetable and Mammal) responses. Three subjects show this effect quite clearly. Subjects 13 and 14 start their bird responses before the second given item, and Subject 19 completes throughout by an alternation of the two given topics.

The effect of given structure is seen again if the distribution of responses belonging to the same semantic class as the given item is considered in relation to their separation from the given item. In Table 2 and Figs. 1 and 2 a tendency is seen for the percentage of responses corresponding to the given topic to decrease with number of spaces from the given item. Values for pre- and post-change responses are recorded separately, and it should be noted that while the former are based on responses to the same initial letters, the latter must refer to different initial letters in the original

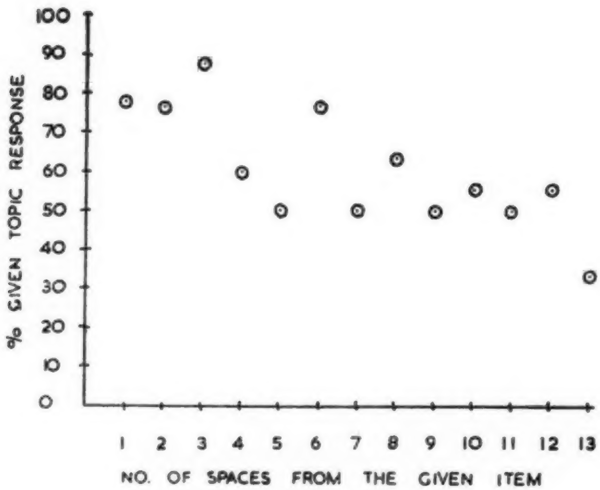


Fig. 1. Test 1. Percentage given topic response by separation from given item. Pre-change responses. (Same initial letters.)

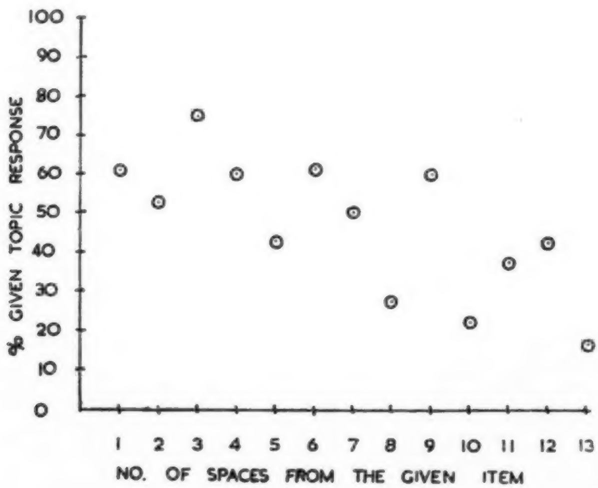


Fig. 2. Test 1. Percentage given topic response by separation from given item. Post-change responses. (Different initial letters.)

TABLE 2

PRE-CHANGE (*Vegetable*) No. of Spaces	% Given Topic Response	POST-CHANGE (*Bird*) No. of Spaces	% Given Topic Response
1	77·8	1	61·1
2	76·5	2	52·9
3	87·5	3	75·0
4	60·0	4	60·0
5	50·0	5	42·9
6	76·9	6	61·5
7	50·0	7	50·0
8	63·6	8	27·3
9	50·0	9	60·0
10	55·6	10	22·2
11	50·0	11	37·5
12	55·6	12	42·9
13	33·3	13	16·7

The percentage of given topic responses according to separation from given items.

list. Although the trend shown is the same, there is as a result a greater degree of scatter in the post-change values. Values are taken only as far as space 13 because of the paucity of cases thereafter. Thus there is a greater chance of completion in terms of the given topic when the number of required pre- or post-change responses is small.

We return now to the question of completion strategies. The response matrix given as Table 1 shows an odd feature which could be coincidence, artefact, or perhaps a point of some significance. Two clusters of given topic responses stand out clearly from the other responses. They are the pre-change responses of Subjects 5 to 8, and the post-change responses of Subjects 11 to 14. In each case, all the spaces between first and second given item, or, where the post-change responses are concerned, all the spaces from the second given item to the end of the list, have been filled with words in the same semantic class as the preceding given item. These unbroken given topic blocks coincide with a range of from four to eight spaces. The only other instances of unbroken topic runs came from the pre-change responses of subjects 11 and 17 (ten and sixteen spaces respectively) and the post-change responses of subject 8 (eleven spaces). In all the other cases permitting longer topic runs the opportunity was not taken. Instead the strategy adopted often led to short topic runs of a similar length to the given topic runs supplied by subjects having a shorter range of available spaces. Thus subject 5, with fourteen spaces for his post-change responses, completed with four birds, five mammals, and five birds. Subject 13 with twelve pre-change response spaces listed eight vegetables and four birds.

The impression arises not only that given structure affects completion strategy, but that the resultant strategies tend to produce a common outcome. If grouping is involved at all, the trend of the classification is towards the supply of short topic runs falling within the favoured range of about four to ten items. These limits are to a

certain extent arbitrary, but their appropriateness is further suggested by the observations :—

(a) That out of sixteen opportunities for making unbroken runs of more than ten items (pre- and post-change responses are considered as contributing independently to the total) only two were actually taken. By contrast, nine of the fourteen potential shorter runs (four to ten spaces) were actually made.

(b) As the unbroken runs made consisted only of responses corresponding to the topic of the previous given item, the preferred length should be indicated by the range of available spaces corresponding to a minimum of responses outside the given topic. The occurrence of such responses can be seen under the heading of " non-given " in the breakdown of results shown below :

	PRE-CHANGE RESPONSES		POST-CHANGE RESPONSES	
Spaces	No. of responses	No. of non-given	No. of responses	No. of non-given
0 - 3	6	4	6	1
4 - 10	49	12	49	16
11+	116	45	116	70

giving a percentage incidence of responses outside the given topic as follows:

SPACES	% NON-GIVEN RESPONSES
0 - 3	41·7
4 - 10	28·6
11+	49·6

Figure 3 shows the same feature the other way round, with pre- and post-change responses combined and a finer breakdown. Here the percentage *given* topic responses have a clear maximum at five to seven spaces.

(c) An attempt at analysis of the completion strategies adopted by the nineteen subjects produces the following results :—

STRATEGY	NO. OF CASES
Sequential associations	3
Alternations and number schemes	4
Groupings based on two or three common topic items	2
Groupings based on four to ten common topic items	8
Groupings based on eleven plus common topic items	2

Further discussion of the point will be witheld until after the description of Test 2. What relevance, if any, this trend in completion strategy has to the previously described tendency to attribute the semantic character of about four to ten responses to a fifty item intelligibility list remains very doubtful. The exploratory test just described is only the beginning of a series designed to clarify the observation, and is in no way critical. It is simply noted that listeners often deduced the semantic unity of a word list from a sub-set of about seven common responses, and that here, where the structure and content of the list are largely the work of the subject, a preference is shown for sub-sets of the same length.

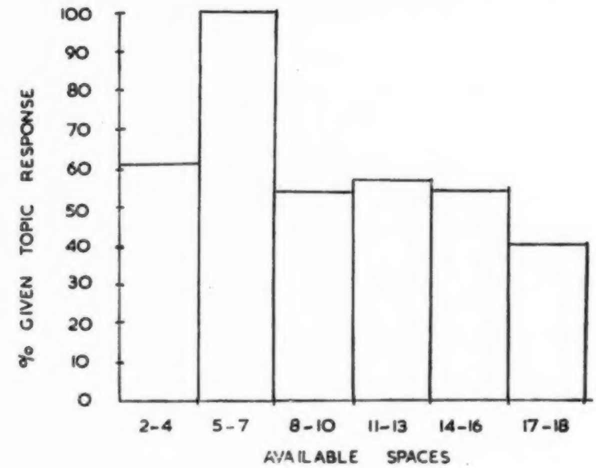


Fig. 3. Test 1. Percentage given topic response by number of spaces available.

TABLE 3

Subjects

Items	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	**V**	**V**	**V**	**V**	**V**	**V**	**V**	**V**	**V**	**V**	**V**	**V**	**V**	**V**	**V**	**V**	**V**	**V**
2	**B**	V	V	V	B	V	V	V	V	V	V	V	V	V	B	V	B	V
3	B	**B**	V	V	M	V	V	V	V	B	V	V	V	V	M	V	V	V
4	V	B	**B**	V	M	V	V	V	V	B	V	V	V	V	B	V	B	V
5	B	V	M	**B**	V	V	V	V	V	V	V	V	V	B	V	M	V	V
6	V	V	B	B	**B**	B	V	V	V	V	V	V	V	B	M	M	B	V
7	V	B	M	B	B	**B**	V	V	B	B	V	V	V	V	M	M	V	V
8	B	B	B	B	M	B	**B**	V	B	B	V	V	V	V	B	M	B	V
9	V	V	B	V	V	B	V	**B**	B	V	V	V	V	B	V	M	V	V
10	V	V	M	V	B	B	B	B	**B**	V	V	V	V	B	V	M	B	M
11	B	B	M	V	B	M	V	B	B	**B**	V	V	V	B	M	M	V	M
12	B	B	M	V	M	M	V	B	B	B	**B**	B	V	B	B	M	B	M
13	V	V	B	B	M	M	V	B	B	V	B	**B**	V	B	V	M	V	B
14	V	V	B	B	V	M	V	B	B	V	M	B	**B**	B	B	M	B	B
15	B	B	B	B	B	M	V	B	B	B	B	B	B	**B**	M	M	V	B
16	B	B	V	B	B	B	V	B	B	B	B	B	B	V	**B**	M	B	B
17	V	V	V	V	M	B	V	B	B	V	B	B	B	B	V	**B**	V	B
18	V	V	V	V	V	B	V	B	B	V	M	B	B	B	M	B	**B**	B
19	B	B	B	V	V	B	B	B	B	B	M	B	B	B	M	B	V	**B**
20	**B**	**B**	**B**	**B**	**B**	**B**	**B**	**B**	**B**	**B**	**B**	**B**	**B**	**B**	**B**	**B**	**B**	**B**

V=Vegetable; B=Bird; M=Mammal. Given Items in **bold type**

Subjects' completions in Test 2

Test 2. Given—One Vegetable, Two Birds

Table 3 lists responses, by topic, for the eighteen subjects in this test.

The matrix shows a similar pattern to that described for Test 1, i.e. representing sequential associations, topic alternations, or number schemes when the first two given items are separated by a minimum or maximum of spaces; runs of given topic responses when the spacing is intermediate. The presence of the third given item, a second bird name, has not then affected the broad pattern of completion strategy. However, the percentage representation of topics is altered, and in line with expectation. In all, three hundred and six responses are involved in Test 2. The change in distribution is as follows :—

TOPIC	TEST 1	TEST 2
Vegetable :	43·9%	46·1%
Bird :	37·4%	40·5%
Mammal :	18·7%	13·4%

Analysis of pre- and post-change responses separately shows the following distribution of topics for the two test conditions :

	PRE-CHANGE RESPONSES		POST-CHANGE RESPONSES	
	Test 1	Test 2	Test 1	Test 2
Vegetable :	64·3%	61·4%	23·4%	30·7%
Bird :	25·7%	24·2%	49·1%	56·9%
Mammal :	9·9%	14·4%	27·5%	12·4%

It will be seen that the over-all differences in topic distribution owe their existence to classificatory changes in the post-change responses, that is to subjects' completion of the spaces between the two given bird items, the region of the list where the difference in test conditions is focussed. The effect of the additional given item on these responses seems to have been :—

(a) To increase the incidence of bird names, so that the representation of this given topic more nearly approaches the representation of the first given topic (Vegetable) in the pre-change responses.

(b) To enhance the tendency to classify in terms of the given (Vegetable and Bird) generally, so that Mammals are poorly represented in the post-change responses as well as in the pre-change responses.

The effect of the third given item is also noticeable when percentage response within the given topic is once more plotted against number of spaces from the given item. As the addition was made in the area of the post-change responses, it could be predicted that the trend observed in Test 1 should remain unaltered for the pre-change responses in Test 2. Table 4 and Fig. 4 show that this expectation was realised. But the comparable graph of the post-change responses (Fig. 5) is very different from that for Test 1. The decrease in percentage response within the given topic now extends only to the halfway point of the scale of space values. A recovery then takes place which must be attributed to the re-inforcing effect of the second given bird item. This is followed

Table 4.

PRE-CHANGE *(Vegetable)* No. of Spaces	% Given Topic Response	POST-CHANGE *(Bird)* No. of Spaces	% Given Topic Response
1	82·4	1	70·6
2	81·3	2	62·5
3	73·3	3	60·0
4	85·7	4	71·4
5	61·5	5	69·2
6	66·7	6	33·3
7	54·5	7	18·2
8	70·0	8	50·0
9	55·6	9	88·9
10	50·0	10	75·0
11	14·3	11	57·1
12	50·0	12	33·3
13	0·0	13	40·0

The percentage of given topic responses according to separation from given items.

by a further decrease, the origin of which is, at the moment, obscure. It could be no more than a reflection of the small number of cases represented by the greater separation values. On the other hand, it is suggestive that this further decline starts from a point where the contributing cases have *ten or more* spaces between the two given bird items in the original list. It is more probably, then, an outcome of the change in completion strategy brought about by the greater separation of given items.

Amongst the completion strategies shown in Test 2, the supply of an unbroken topic run of more than ten items was again an exceptional occurrence. As in Test 1, the preferred strategy where grouping was adopted was to make short topic runs. Thus, Subject 18 with seventeen spaces between the first and second given items completed with eight vegetables, three mammals and six birds. Subject 6 with twelve post-change spaces supplied three birds, five mammals and four birds.

Out of the fourteen opportunities for making unbroken longer runs, only one was actually taken (Subject 13, pre-change responses) compared to seven out of fourteen for the potential shorter runs (four to ten spaces). As in Test 1, all unbroken runs were in terms of the topic of the previous given item. When the occurrence of responses outside the given topic is associated with number of available spaces, a similar picture to that provided by Test 1 emerges :—

	PRE-CHANGE RESPONSES		POST-CHANGE RESPONSES	
Spaces	No. of responses	No. of non-given	No. of responses	No. of non-given
0 - 3	6	0	6	4
4 - 10	49	10	49	8
11 +	98	48	98	54

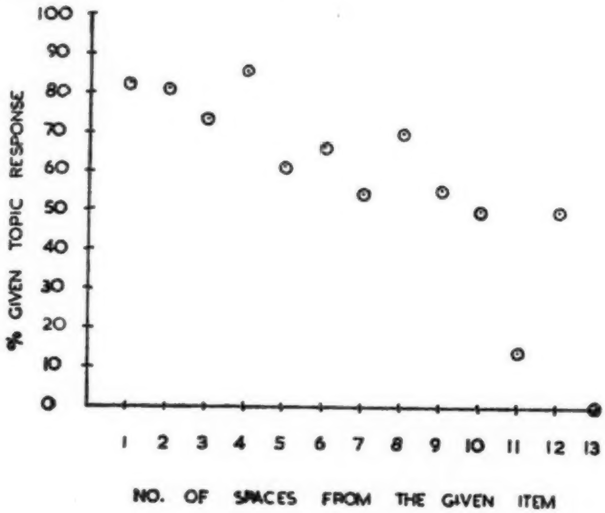


Fig. 4. Test 2. Percentage given topic response by separation from given item. Pre-change responses. (Same initial letters.)

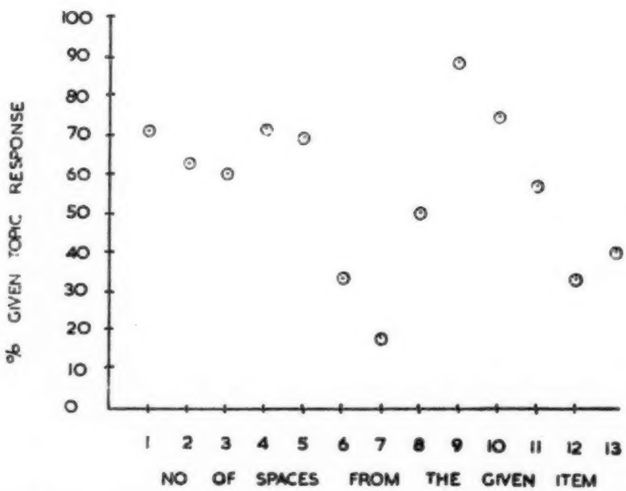


Fig. 5. Test 2. Percentage given topic response by separation from given item. Post-change responses. (Different initial letters.)

Giving a percentage incidence of responses outside the given topic as follows :

SPACES	% NON-GIVEN RESPONSES
0 - 3	33·3
4 - 10	18·4
11+	52·0

Fig. 6 shows that *given* topic responses had their maximum incidence at five to ten spaces in Test 2.

When an analysis of the completion strategies used in this test is attempted, the preference for classification based on short topic runs appears again :—

STRATEGY	NO. OF CASES
Sequential associations	2
Alternations and number schemes	3
Groupings based on two or three common topic items	3
Groupings based on four to ten common topic items	8
Groupings based on eleven plus common topic items	2

Speculation

The expected progression for this experimental series is that classification will pass from the present degree of use of the three alternative topics (Vegetable, Bird and Mammal), through a dropping of the topic which is never given, to a completion exclusively in terms of the predominant given topic. Whether or not such an orderly development is forthcoming, a feature has already appeared which is worth emphasising. This is the tendency, under the conditions of given structure reported here, to complete by runs of common topic responses not exceeding ten items.

There is, incidentally, no incompatibility between the possible continued use of these short runs and completion wholly in terms of one topic, if the latter appears under the conditions hypothesized—i.e., a particular given structure involving an increased number of semantically similar given items. The observation refers to the number of common topic responses *supplied by the subject* before, after, or between given items, and not to the length of the sequence produced when the given items are themselves counted in. The latter are used by the subject as reference points in completion. However, the most telling evidence that this short run feature is not just an artefact must come from conditions where the ratio of number of given items to list length is a minimum and/or the spacing of given items allows the possibility of longer runs. The two tests reported here represent these conditions for a twenty item list, and a combination of their results produces the following evidence :—

(a) The proportion of actual to possible unbroken longer run completions (eleven spaces or more) is 3/30. The proportion of actual to possible unbroken shorter run completions (four - ten spaces) is 16/28.

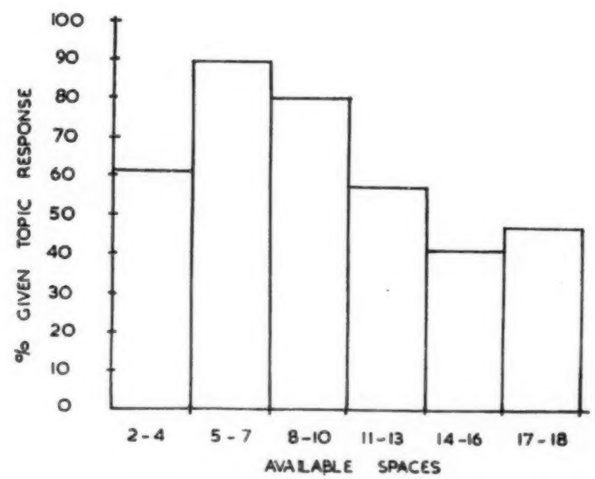


Fig. 6. Test 2. Percentage given topic response by number of spaces available.

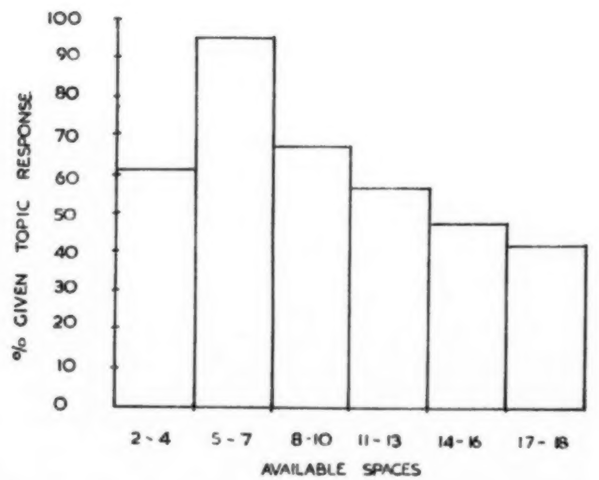


Fig. 7. Tests 1 and 2. Percentage given topic response by number of spaces available.

(b) Associating the occurrence of responses outside the given topic with number of available spaces results in the following percentage distribution :—

SPACES	% NON-GIVEN RESPONSES
0 - 3	37·5
4 - 10	23·5
11+	50·7

(c) The complementary distribution of given topic responses, using a finer scale of space values is :—

SPACES	% GIVEN RESPONSES
2 - 4	61·0
5 - 7	94·5
8 - 10	67·0
11 - 13	56·5
14 - 16	47·5
17 - 18	42·0

These results are shown graphically in Fig. 7.

(d) The frequency distribution of completion strategies employed by the thirty-seven subjects appears thus :

STRATEGY	NO. OF CASES
Sequential associations	5
Alternations and number schemes	7
Groupings based on two or three common topic items	5
Groupings based on four to ten common topic items	16
Groupings based on eleven plus common topic items	4

A genuine preference does seem to exist for a classification based on groups of around seven items. With a range of available spaces approximating this value subjects show a greater tendency to make a simple extrapolation from the semantic class of the given item. When the range of available spaces is in some excess of this value, the other topics are called into response, but, if groups are used, their size tends to remain constant around the same value.

What could be the source of this preference ? Miller (1956) has drawn attention to the ubiquitous presence of values around seven in the data concerning human capacity for processing information. In particular, the immediate memory span for items, irrespective of their information value, can be consistently defined by a length reminiscent of that of the topic runs reported here. The organizational processes which achieve information enrichment of this constant item span are themselves closely allied to the activity of classification. May it be that when the subject has the opportunity to create a classified word list he does so in a way commensurate with his

immediate memory span ? For example, it is a reasonable assumption that the division of seventeen items into eight vegetables, three mammals, and six birds produced for Subject 18 a classification which was mnemonically more acceptable than a list of seventeen vegetables with its attendant order information problem. Some attempt will be made to test the validity of this suggestion as the investigation proceeds and more data becomes available.

References

Bruce, D. J. (1956). Effects of context upon intelligibility of heard speech. Information Theory—Third London Symposium (London), 245.

Foley, J. P. and Macmillan, Z. L. (1943). Mediated generalization and the interpretation of verbal behaviour: *V.* "Free association" as related to differences in professional training. *J. Exp. Psychol.*, 33, 299.

Howes, D. and Osgood, C. E. (1954). On the combination of associative probabilities in linguistic contexts. *Amer. J. Psychol.*, 67, 241.

Jung, C. (1917). Analytical Psychology (London).

Karwoski, T. F. and Berthold, F. (1945). Psychological studies in semantics: *II.* Reliability of free association tests. *J. Soc. Psychol.*, 22, 87.

Karwoski, T. F. and Schachter, J. (1948). Psychological studies in semantics: *III.* Reaction times for similarity and difference. *J. Soc. Psychol.*, 28, 103.

Kent, G. H. and Rosanoff, A. J. (1910). A study of association in insanity. *Amer. J. Insanity*, 67, 37 - 96, 317.

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychol. Rev.*, 63, 81.

A COMPARATIVE STUDY OF TWO HESITATION PHENOMENA

FRIEDA GOLDMAN-EISLER
University College, London

The durations of hesitation devices such as the sounds /ɑ, ɛ, æ, r, ə, m/, also called *filled pauses,* were measured and compared with the durations of silent hesitations or *unfilled pauses.* Their individual consistency and psychological significance were also investigated and the relation to uncertainty of filled pauses and unfilled pauses respectively was compared. It appears that under certain conditions of speech production the two hesitation phenomena reflect different internal processes.

INTRODUCTION

Previous work by the writer on hesitation pauses concentrated on the silences which interrupt speech utterance (Eisler, 1958a, b). A recent paper by Maclay and Osgood on " Hesitation phenomena in Spontaneous English " (1959) deals with four types of hesitation phenomena ; beside the silent pauses which they call *unfilled pauses* (UP) they also studied the occurrences of sounded hesitation devices, i.e., the /ɑ, ɛ, æ, r, ə, m/ sounds of hesitation which they call *filled pauses* (FP), as well as repeats and false starts. We shall here be interested only in the hesitation phenomena of unfilled and filled pauses.

Maclay and Osgood recorded these phenomena by taking counts of their frequency. Thus unfilled pauses are counted by occurrence irrespective of their duration, in the same way as the filled pauses. (The writer's own method of recording unfilled pauses consists, apart from recording their occurrence, in measuring their duration from visual speech recordings (Eisler, 1956, 1958a, b).)

Maclay and Osgood have also taken note of the position of filled and unfilled pauses in the sentence, and in relation to the grammatical function of words.

The following results relevant to the present paper emerged from their analysis:

(1) Both filled pauses and unfilled pauses are found to occur more frequently before lexical words than before function words. But unfilled pauses are relatively more likely to appear before lexical words.

(2) For those constructions that can be analysed statistically, filled pauses occur more frequently at phrase boundaries than within phrases. These are statistically significant tendencies, not cases of absolute complementary distribution in the linguistic sense.

(3) Filled pauses and unfilled pauses were a matter of individual differences ; the relative " preference " for hesitation phenomena of different types seems to be an aspect of individual style of speaking.

It should be noted that conclusion (2) is a corollary of conclusion (1) as phrases commonly start with function words while most lexical words occur within phrases. With the greater uncertainty in the choice of lexical words it follows that unfilled pauses are better indicators of uncertainty of choice than filled pauses.

Maclay and Osgood suggest that the distinction between filled and unfilled pauses as indicated in (1) and (2) lies mainly in the duration of the non-speech interval. They write:

"Let us assume that the speaker is motivated to keep control of the conversational 'ball' until he has achieved some sense of completion. He has learned that unfilled intervals of sufficient length are the points at which he has usually lost his control—someone else has leapt into the gap. Therefore, if he pauses long enough to receive the cue of his own silence, he will produce some kind of signal (ah, m, er) or perhaps a repetition of the immediately preceding unit, which says in effect: 'I'm still in control—don't interrupt me'. We would thus expect filled pauses and repeats to occur just before points of highest uncertainty, points where choices are most difficult and complicated.... This assumption that 'ah' type pauses are reactions of the speaker to his own prolonged silences at points of difficult decision is consistent with our finding that these two pause-types are merely statistically, not absolutely, different in distribution.... The less probable the sequence, the more prolonged the non-speech interval and hence the greater the tendency for an 'ah' or a repetition."

Maclay and Osgood's suggestion concerning unfilled pauses and their relation to difficult decisions is in keeping with the writer's own experimental results as was pointed out by these authors (Maclay and Osgood, 1959) and further evidence derived from measurements of pause duration has since been produced by the writer (1961) to demonstrate that "the less probable a sequence the more prolonged the non-speech interval".

Maclay and Osgood's observations on the distinction between filled and unfilled pauses raising the question of the relative significance of the former has stimulated the present investigation. Its purpose has been to see to what extent the introduction of the criterion of time might help to illuminate further the relative functions of filled and unfilled pauses.

Material

The speech samples used for this investigation were taken from an experiment, reported elsewhere (Eisler, 1961), which was concerned with the relation of hesitation pauses to degree and level of selection and uncertainty. It consisted in showing subjects cartoon stories without captions (of the kind regularly published in the "New Yorker" magazine) asking them first to describe the content of the stories and then to formulate the meaning, point, or moral of the story. Experimental conditions were thus created for the study of pauses (a) in speech produced within a relatively concrete situation, i.e., a given sequence of events (through their description) and (b) in speech uttered in the process of abstracting and generalising from such events (through summarising their meaning).

The speech produced by the subjects was recorded, transcribed and visual records obtained of the sequences of sound and silence, the length of which were measured, as described in a previous paper (Eisler, 1956).

TABLE 1

	DESCRIPTIONS Cartoons		SUMMARIES Cartoons	
Subjects	1	2	1	2
Ha	5.88%	0.00%	0.00%	2.45%
Tr	0.62	2.56	0.88	2.35
Co	4.80	4.26	15.38	7.14
Sa	1.23	0.88	10.34	0.00
Gi	5.26	2.21	9.01	5.45
Ne	3.48	5.81	18.52	2.78
Am	4.63	1.06	1.47	4.21
Do	0.00	0.00	57.10	1.33

Percentage of total pause time taken up by filled pauses.

The results showed that (a) speech describing observed events contains considerably less hesitation (as measured by duration of pauses) than speech produced in conveying the meaning of these events.

(b) Hesitancy (pause length per speech unit) which is independent of the length of utterances in descriptive speech, becomes a function of brevity of verbal expression when the meaning of the cartoon stories is summarised. Greater conciseness in summarising was associated with more hesitation.

(c) A transitional analysis executed on descriptive speech and summaries separately showed the summaries to carry words of significantly greater uncertainty than the descriptions. (Oral communication at 4th London Symposium on Information Theory, 1960, to be published.) The material of this experiment was used for the present study.

Measurements were taken of the durations of the filled as well as the unfilled pauses ; this was done for descriptions and summaries separately.

RESULTS

1. *Relative length of filled and unfilled pauses.*

Time measurements of the filled pauses which occurred in the speech of nine subjects, describing and summarising 7 - 9 cartoons each, showed that the duration of filled pauses ranged between 0·2 to 0·8 sec. each. The total length of time taken up by filled pauses in relation to the total non-speech pauses (filled plus unfilled pauses) covered a range from 0·0 to 18·5% of the total pause time with a single stray value of 57·1% where there was very little pausing of any kind and the verbal statement itself was very short. The mean percentage of the total pause time taken up by filled pauses was 5·7% (including the value of 57·1% in the total). This figure, however, covers a very wide spread (see Table 1) with nearly two-thirds of the filled pauses taking up less than 5% of the total pause time, and three-quarters less than 6%.

Relating the length of filled pauses (FP) and unfilled pauses (UP) to the output of speech (number of words produced), in the descriptions FP time per word produced

TABLE 2

Subjects	DESCRIPTIONS FP/w	UP/w	SUMMARIES FP/w	UP/w
Ha	0.00 sec.	0.53 sec.	0.01 sec.	1.69 sec.
Tr	0.02	0.56	0.04	1.88
Wi	0.01	0.30	0.02	1.00
Sa	0.01	0.33	0.01	1.62
Gi	0.02	0.27	0.01	0.46
Ne	0.01	0.38	0.04	0.72
An	0.09	0.40	0.50	0.74
Do	0.00	0.30	0.01	0.77
Th	0.01	0.22	0.01	0.61

Time occupied by filled pauses (FP/w) and unfilled pauses (UP/w) in descriptions and in summaries, expressed in seconds per word produced: mean values based on 7 - 9 cartoon experiments with each subject.

TABLE 3

Subjects	DESCRIPTIONS (FP, UPt)	SUMMARIES (FP, UPt)
Tr	0.928**	0.667*
Sa	0.867**	—
Gi	0.372	0.596
An	0.955**	—

** Significant at 1% level.
* Significant at 5% level.

Rank correlations between filled pause frequencies (FP) and the total time occupied by unfilled pauses (UPt).

(FP/w) was 0·013 sec. against UP time per word (UP/w) of 0·365 sec. and in the summaries, 0·023 sec. against 1·054 sec.

2. *Filled pauses, unfilled pauses and points of uncertainty.*

The relation was studied of the frequency of filled pauses to the total time of unfilled pauses (FP/UPt) in descriptions and those summaries which were long enough to permit such correlation, for each of four subjects separately. (The correlations were based on nine and seven cartoons for each of two subjects.) The frequency and variability of filled pauses for the rest of the subjects were too low to justify correlation. Table 3 shows that for three out of the four subjects the frequency of the filled pauses is a function of the duration of the unfilled pauses. It has been shown (Eisler, 1961) that the total length of unfilled pauses is a function of the total length of verbal productions. The longer we speak, the more words we produce and the more time we spend being silent. As linguistic and speech phenomena are functions of time, it is

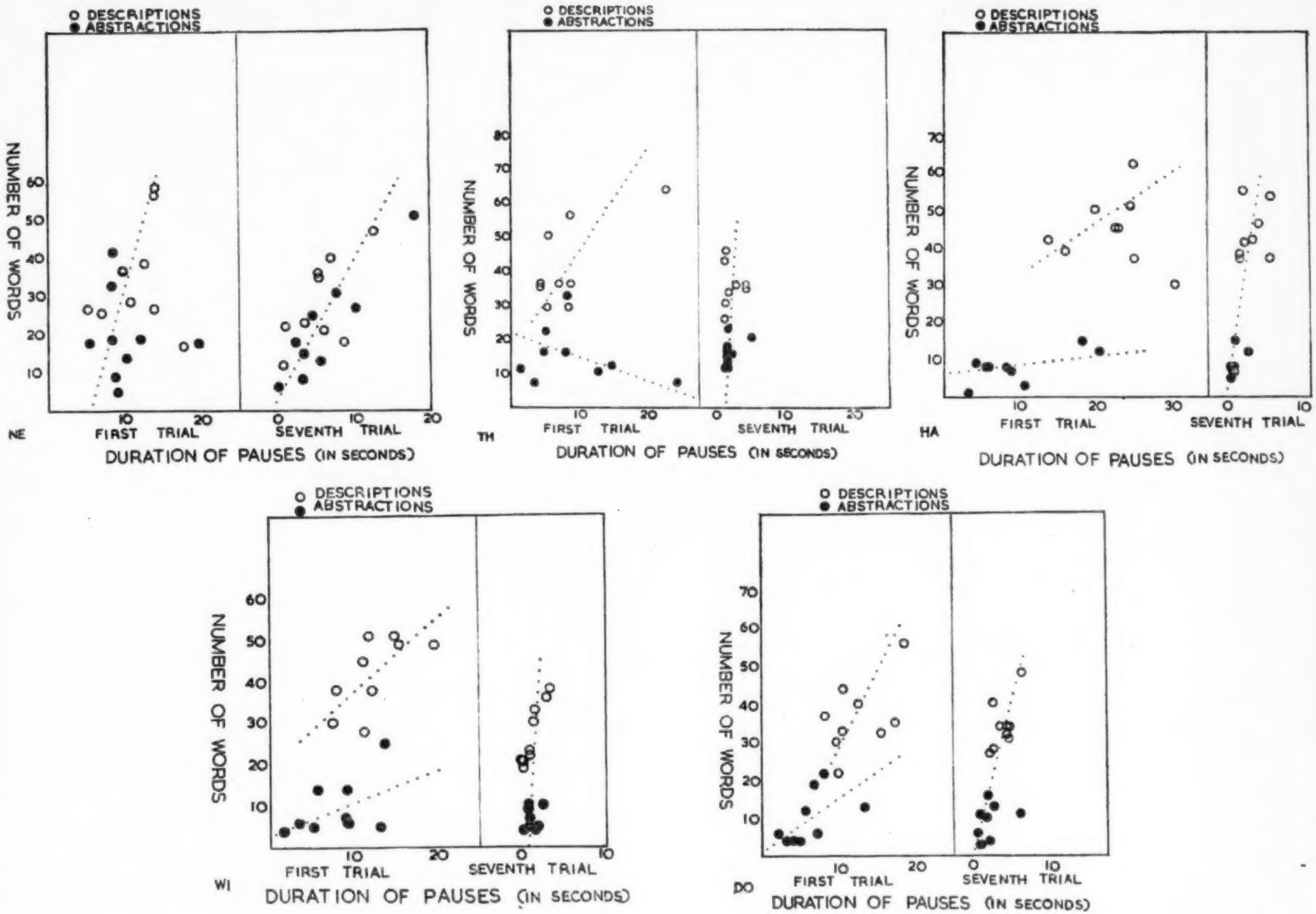


Fig. 1. The rate of growth of unfilled pause time with total speaking time, measured by the number of words produced, for 9 different subjects. Each section shows the results for four verbal tasks: spontaneously describing pictures and abstracting their meaning at a first trial, and repeating

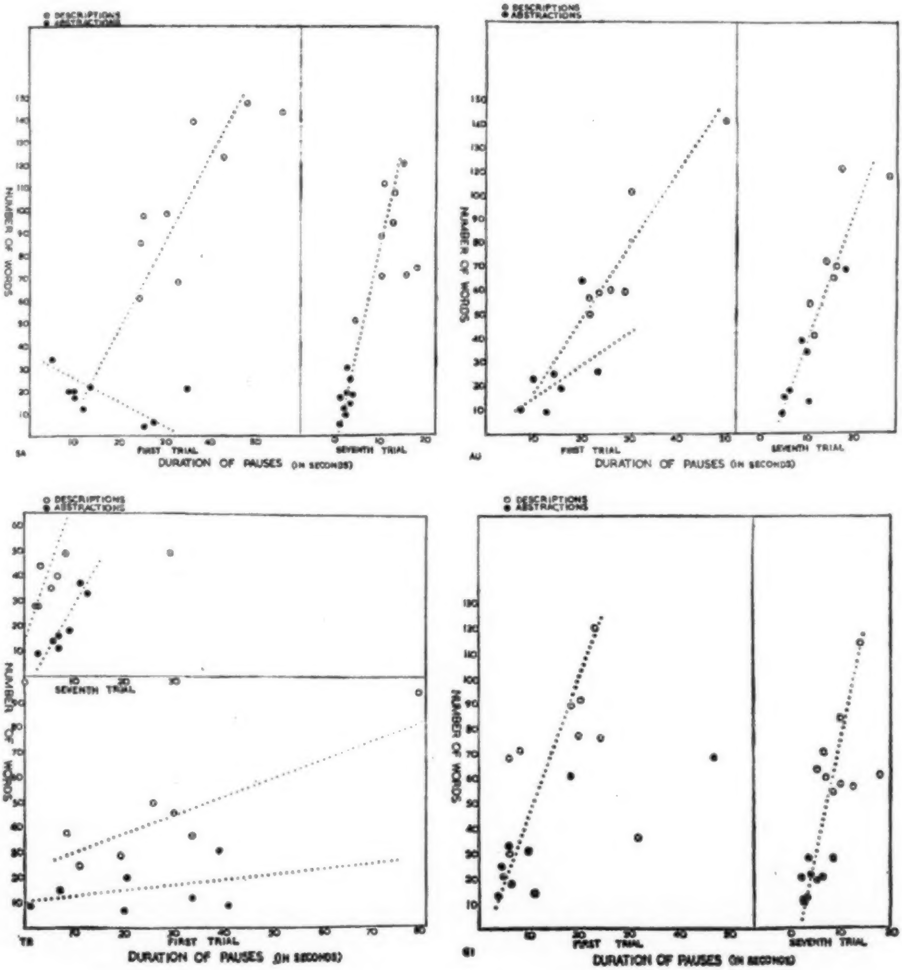


Fig. 1 (cont.).

not surprising to find that the frequency of filled pauses increases with the increasing total length of unfilled pauses.

Fig. 1 shows however that the rate of growth of unfilled pause time with total speaking time differs for different individuals, and in the same way we must expect that the rate of increase of filled pauses relative to unfilled pause time will be a discriminating factor in different individuals, and under different conditions.

Fig. 1. The rate of growth of unfilled pause time with total speaking time, measured by the number of words produced, for 9 different subjects. Each section shows the results for four verbal tasks: spontaneously describing pictures and abstracting their meaning at a first trial, and repeating

TABLE 4

Subjects	DESCRIPTIONS FP/UPt	SUMMARIES FP/UPt
Ha	1 : 77	1 : 83
Tr	1 : 13	1 : 14
Co	1 : 19	1 : 13
Sa	1 : 7	1 : 16
Gi	1 : 4	1 : 6
Ne	1 : 7	1 : 4
Au	1 : 4	1 : 4
Do	1 : 114	1 : 19
Th	1 : 9	1 : 10

Ratios of filled pause occurrence to the time occupied by unfilled pauses (in seconds).

TABLE 5

Subjects	Descriptions	Summaries
Ha	0.013	0.012
Tr	0.079	0.074
Wi	0.052	0.079
Sa	0.138	0.062
Gi	0.249	0.175
Ne	0.140	0.250
An	0.264	0.273
Do	0.009	0.005

Mean FP/UPt rates.

Table 4 shows the ratio for nine subjects of the frequency of filled pauses to the duration of unfilled pauses (in seconds) for descriptions and summaries. It illustrates the considerable differences among individuals in the silence they can tolerate without breaking it with vocal activity. The exceptional ratio 1 : 114 for subject Do. must however be interpreted in the light of the fact that this subject was particularly curt in utterance and short in pausing. The infrequency of filled pauses under these circumstances falls in well with Maclay's and Osgood's suggestion that " ah " or " m " sounds are speakers' reactions to their own prolonged silences. The silences of subject Do. were rarely long enough to stimulate him into signalising vocally that he was still talking. For the rest, the consistency of individual ratios of filled pause frequency to unfilled pause duration is evident even from these average figures when we compare descriptions and summaries. An analysis of variance based on filled pause occurrence per second of unfilled pause time for six subjects shows the degree and significance of this consistency (Table 6).

A coefficient of reliability calculated from the variance ratio was 0·950. It is also evident from this analysis of variance that the different levels of speech production

TABLE 6

SOURCE	SUM OF SQUARES	df	VARIANCE ESTIMATE
Descriptions and summaries	0.0172	1	0.0172
Between subjects	0.5226	5	0.1045
Interactions	0.5969	5	0.1195
Within subjects (error)	0.6041	116	0.0052

Between subjects/Error	$F = 20.1$	$p < 0.001$
Descriptions and summaries/Error	$F = 3.3$	Not significant
Interaction	$F = 22.9$	$p < 0.001$

Analysis of variance: filled pause rate (FP/UPt) for descriptions and summaries, based on 6 subjects and 128 cartoons (5 cartoons for 4 subjects and 7 cartoons for 2 subjects).

operating in descriptions and summaries which were reflected most significantly in the length of unfilled pauses (Eisler, 1961) showed no systematic effect on the rate of filled pause occurrence per second of unfilled pauses. However, individual differences accounted for only about half of the variance whilst the other significant half was due to interaction effects between subjects and cartoons.

Conclusions and discussion

Three conclusions seem to be justified on the basis of the above results:

(a) That the ratio of filled pause occurrence to the time occupied by unfilled pauses can be classed as a speech habit characteristic of individuals.

(b) That in contrast to silent hesitation (unfilled pauses) which have also been shown to contain a habitual factor (Eisler, 1961) deviations from habitual filled pause rate are not stimulated by cognitive factors such as degree of abstraction in speech production or difficulty of choice as measured by transition probability.

(c) That on the other hand, judging by the significant interaction between subjects and cartoons, factors connected with the content of the cartoons do seem to stimulate subjects to deviate from their habitual filled pause rate. This suggests an emotional factor.

The two hesitation phenomena of filled and unfilled pauses would thus appear to reflect different internal processes, cognitive activity being accompanied by an arrest of external activity (speech or non-linguistic vocal action) for periods proportionate to the difficulty of the cognitive task, while emotional attitudes would be reflected in vocal activity of instantaneous or explosive nature.

This interpretation was put to the test by correlating the mean filled pause rate (FP/UPt) for nine subjects with their mean hesitancy (unfilled pause length per word, P/w). The correlation Spearman's r, was $-0{\cdot}665$, significant at the 0·05 level of probability for the summaries. There was no significant relation ($r = 0{\cdot}100$) for the descriptions, but this might have been expected from the small range of individual differences in hesitancy (P/w) which was not only less in extent, but also less discriminating between individuals in the descriptions. The summaries which represent responses to a considerably more difficult cognitive task resulted not only in greater hesitancy generally,

TABLE 7

Subjects	DESCRIPTIONS P/w	SUMMARIES P/w
Th	0.22 sec.	0.61 sec.
Ha	0.53	1.69
Tr	0.56	1.88
Wi	0.30	1.00
Sa	0.33	1.62
Gi	0.27	0.46
Ne	0.38	0.72
Au	0.40	0.74
Do	0.30	0.77

Pause time per word produced in descriptions and in summaries.

but also in wider differentiation between the specific hesitancy of individuals, as may be seen from Table 7. The negative correlation of these latter values (P/w) with the subject's mean filled pause rate (FP/UPt) shows that subjects whose hesitancy in formulating summaries was greater, were less inclined to break their silences with " ah " or " m " sounds, while subjects whose silent pauses were shorter, uttered more of such sounds. This would seem to contradict Maclay and Osgood's suggestion that filled pauses are responses to length of unfilled pauses, but it is a conclusion applicable under specific conditions, namely in a situation requiring high level cognitive activity. Under conditions requiring processes of abstraction and generalisation those who hesitated longer in silence, who also produced more concise statements and words which were less predictable, produced fewer filled pauses per second of unfilled pause time, while the less hesitant subjects who produced the more long-winded summaries and more predictable words produced filled pauses at shorter intervals of silence.

Thus those who consistently achieved superior (more concise) stylistic and less probable linguistic formulations are consistently inclined towards delay of action and tolerance of silence, whilst the inferior stylistic achievement (long-winded statement) of greater predictability is linked to greater verbal as well as vocal activity and to intolerance of silence.

REFERENCES

GOLDMAN-EISLER, F. (1956). The determinants of the rate of speech output and their mutual relations. *J. Psychosom. Res.*, 1, 137.

GOLDMAN-EISLER, F. (1958a). Speech production and the predictability of words in context. *Quart. J. exper. Psychol.*, 10, 96.

GOLDMAN-EISLER, F. (1958b). The predictability of words in context and the length of pauses in speech. *Language and Speech*, 1, 226.

GOLDMAN-EISLER, F. (1961). Hesitation and information in speech. In *Proceedings of the 4th London Symposium on Information Theory* (in the press).

MACLAY, H. and OSGOOD, C. E. (1959). Hesitation phenomena in spontaneous English speech. *Word*, 15, 19.

SEGMENT INVENTORIES FOR SPEECH SYNTHESIS *

EVA SIVERTSEN
University of Michigan and Norges Lærerhøgskole, Trondheim

Speech synthesis may be based on a segmentation of the speech continuum either into simultaneous components or into successive time segments. The time segments may be of varying size and type: phonemes, phoneme dyads, syllable nuclei and margins, half-syllables, syllables, syllable dyads, and words.

In order to obtain an estimate of the size of the segment inventory for each type of segment, a phonological study was made of the particular phoneme sequences which occur in English, particularly in relation to the immediate constituents of the syllable (nucleus and margin) and to the syllable. An estimate was also made of the number of prosodic conditions required for each type of phoneme sequence.

It was found that in general there is a direct relationship between the length of the segment and the size of the inventory. However, when the borders of the proposed segments do not coincide with the borders of linguistic units, the inventory has to be relatively large.

The value of using the various types of segment for speech synthesis is discussed, both for basic research on speech and for practical application to a communication system with high intelligibility.

1. INTRODUCTION

1.1. *Types of Synthesis*

There are basically two methods for synthesizing speech, depending on the underlying type of segmentation of the speech continuum.

(a) If the speech continuum is segmented into simultaneous components, speech may be synthesized by controlling the various parameters independently and simultaneously. The parameters may be physiological, such as larynx activity, nasalization, point of oral constriction, degree of oral constriction, and lip opening ; or they may be acoustical, such as fundamental frequency, laryngeal spectrum, formant frequencies, formant bandwidths, formant amplitudes, and overall amplitude.

(b) If the speech continuum is segmented in time, speech may be synthesized from successive building blocks. The building blocks may either be taken from normal recorded utterances, or they may be produced electronically.

* *This work was supported by the Information Systems Branch of the Office of Naval Research of the United States Navy, under contract Nonr 1224(22), NR 049-122, and by the Norwegian Advanced Teachers' College (Norges Lærerhøgskole), Trondheim, Norway.*

This paper is based on the material which appears in Report No. 5 from the Speech Research Laboratory of the University of Michigan, henceforth called SRL Report No. 5. *The appendices of the Report contain some material which is excluded here. Thus, the occurrence of particular phoneme combinations is specified for each of the major sources used for this study. In the present paper only an overall list for English in general is given.*

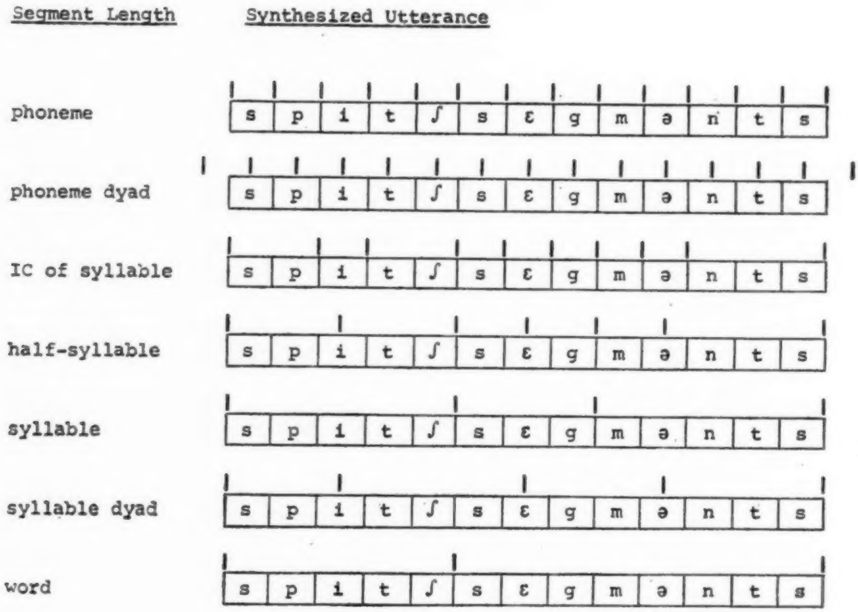


Fig. 1.

In the latter type of synthesis, the building blocks may be of varying size (Peterson and Sivertsen, 1960). They may be phonemes, phoneme dyads, immediate constituents of the syllable, half-syllables, syllables, syllable dyads, or words. Fig. 1 shows which particular segments, of each type, would be needed for synthesizing the utterance /spitʃ sɛgmənts/ *speech segments*. One might expect that the number of segments required for storage will increase with the length of the building blocks.

1.2. *Phonemes*

It would seem that if the phoneme is chosen as the basic unit, a very limited set of segments would be required. Thus, for one type of American English one would need 15 syllable nuclei and 22 consonants, multiplied by the number of prosodic conditions required for each phoneme. Assuming that we need three pitch levels and four pitch glides, and that the glides occur on the nuclei (cf. 2·33, 3·3), the following number of segments would be required:

Voiceless consonants	8×1	$= 8$
Voiced consonants	14×3	$= 42$
Syllable nuclei	15×7	$= 105$
Total		$= 155$

However, speech synthesized in this way would hardly be intelligible, even if each phoneme was represented by several allophones in the segment inventory, one for each general type of phonological context. One reason is that the target positions of the phonemes are not the only clues for phoneme recognition: the transitions between them may be equally important. Besides, these target values are convenient and necessary abstractions set up by linguists and phoneticians, but they are not always actually present in the speech wave. It seems necessary to include more of the dynamics of speech in the segments. The phoneme-segment approach has been studied by Harris (1953).

An alternative would be to employ segments of less than phoneme length. This method has not been tried, except as part of a time-frequency compression-expansion system (Fairbanks, Everitt, and Jaeger, 1954). It is not immediately apparent that such an approach would be useful for speech synthesis, and no estimates are available for the number of segments required.

1.3. *Phoneme Dyads*

The Speech Research Laboratory of the University of Michigan has examined the possibility of using the phoneme dyad as the basic segment for speech synthesis in order to account for some of the dynamics of speech (Peterson, Wang, and Sivertsen, 1958). This approach has been tested only by mechanically cutting out segments of magnetic tape from recordings of natural utterances, and resplicing them, but it could conceivably also be used in electronic synthesis. The dyad has as its centre the transition between two phonemic or phonetic units, the cut being made in the relatively sustained part of the units. Thus [kæt] would be made up of the segments

$$\frac{\#}{2}\frac{\text{k}}{2} + \frac{\text{k}}{2}\frac{\text{æ}}{2} + \frac{\text{æ}}{2}\frac{\text{t}}{2} + \frac{\text{t}}{2}\frac{\#}{2}$$

In general, n + 1 segments are required to synthesize a sequence of n phonemes.

If each syllable nucleus and each consonant is represented by one allophone in the list of essential phonetic units, in one type of American English the dyads would be combinations of the following units:

Syllable nuclei	15
Consonants	22
Juncture, or silence	1
Total	38

If one assumes that all combinations occur, the total number of dyads would be 38 × 38 = 1444. However, it is not likely that all combinations will occur. On the other hand, some phonemic units may have to be represented by more than one allophone in the list of basic units. Finally, the number of dyads would have to be multiplied by several conditions of prosody.

Wang and Peterson (1958) made an estimate of the total number of phoneme dyad segments required for speech synthesis. They first established an inventory of essential phonetic units, including silence, syllabic consonants, [tʃ] and [dʒ] as monophonemic units, and two allophones of each of the phonemes /ə, p, t, k/, and considering [aɪ,

aʊ, ɔɪ] sequences of vowel + glide (/j/ or /w/). Their total inventory is 43. They then examined which particular two-member combinations would be necessary for speech synthesis. Computations based on their Figs. 1 and 2 show that an inventory of 1218 dyads is sufficient, when intonation differences are ignored. In order to account for intonation, 8460 dyad segments are required.

1.4. *Longer Segments*

In the following pages some of the other possibilities of segment types will be explored. The size of the segment inventory in each case will be estimated, and for the two shortest segment types the particular phoneme sequences which would be needed will be listed.

2. Data

2.1. *Sources*

There are a number of studies of phoneme sequences in English. Most phonological descriptions will include a statement of the distribution of the various phonemes, the restrictions on their occurrence, and the syllable structure. There are also a number of studies of the frequency of occurrence of phonemes, syllables, and words. The data which form the basis of the present paper were taken from those studies which state the distribution of phonemes most completely and systematically. An attempt was made to establish which particular phoneme sequences occur in their lists or are possible according to their formulæ or rules of distribution.

The following publications were excerpted as fully as possible:

G. Dewey, Relativ Frequency of English Speech Sounds.

N. R. French, C. W. Carter, Jr., and W. Koenig, Jr., The words and sounds of telephone conversations.

K. Malone, The phonemic structure of English monosyllables.

J. D. O'Connor and J. L. M. Trim, Vowel, consonant, and syllable—a phonological definition.

B. Trnka, A Phonological Analysis of Present-Day Standard English.

B. J. Wallace, A Quantitative Analysis of Consonant Clusters in Present-Day English.

B. L. Whorf, Linguistics as an exact science.

In addition, a number of other phonological studies of English were consulted, and the lists were supplemented by material which the author has otherwise found. The following publications were of particular importance:

L. Bloomfield, Language.

A. A. Hill, Introduction to Linguistic Structures.

D. Jones, An English Pronouncing Dictionary.

J. S. Kenyon and T. A. Knott, A Pronouncing Dictionary of American English.

C. Wood, Wood's Unabridged Rhyming Dictionary.

The two pronouncing dictionaries were used for an unsystematic check on any additional word-initial phoneme sequences, and the rhyming dictionary provided data

TABLE 1

PUBLICATION	DIALECT	CORPUS
G. Dewey, Relativ Frequency of English Speech Sounds.	American English.	1,370 most common syllables, making up 93·3% of the total, in 100,000 words of printed material, transcribed according to *Funk and Wagnall's Standard Dictionary*.
N. R. French, C. W. Carter, and W. Koenig, Jr., The words and sounds of telephone conversations.	American English. (various dialects).	737 most common words, making up 96% of the total occurrence of words, in 80,000 words of telephone conversations.
K. Malone, The phonemic structure of English monosyllables.	American English (various dialects).	All monosyllabic words.
J. D. O'Connor and J. L. M. Trim, Vowel, consonant, and syllable—a phonological definition.	Received Pronunciation of Southern England.	All monosyllabic words, excluding proper names, foreign, learned, and rare words.
B. Trnka, A Phonological Analysis of Present-Day Standard English.	Received Pronunciation of Southern England.	All monomorphemic words listed in *An English Pronouncing Dictionary* by D. Jones.
B. J. Wallace, A Quantitative Analysis of Consonant Clusters in Present-Day English.	American English	10,000 words of tape-recorded running conversation in a Mid-Western university community.
B. L. Whorf, Linguistics as an exact science.	" Standard Mid-Western American."	" The monosyllabic word in English."

List of publications excerpted, with a view to establishing the phoneme sequences occurring in English.

on final sequences. A similar check was not undertaken for word-medial position.

The number of phoneme sequences which are listed in or which are considered possible according to each of these studies varies, partly because they deal with different dialects of English, partly because they are based on corpora which differ in type and size, and partly because they represent different phonemic analyses of English. In order that one may evaluate and compare the figures computed from their data, a brief description of the corpus of each study is given below, with a discussion of the interpretation of their material. Table 1 gives the descriptions in condensed survey form.

Dewey's overall data comprise 100,000 words of written material, running texts, which he transcribes according to Funk and Wagnalls' *Standard Dictionary*. In this material he finds 4,400 different syllables. Out of these, 1,370 syllables occur more than 10 times each, and make up 93·3% of the total material. In our tables (3.1, 4.1)

under " Dewey I " are listed the 1,370 most common syllables, whereas " Dewey II " represents the expanded list given in his Appendix B, where he lists all those syllable-initials and syllable-finals he has ever found. Dewey II does not give any information about the combinatory possibilities between syllable nuclei and margins.

He does not discuss how he divides the words into syllables, but it is assumed that he follows the principles laid down in the dictionary. It is likely that much of what Dewey lists as syllable-initial and syllable-final under another type of analysis could be considered parts of interludes (Hockett, 1955 and 1958).

Syllabic consonant clusters (22 in Dewey I, 51 in Dewey II), with /n/ or /l/ as the syllabic centre, are not included in the present data.

French, Carter, and Koenig, henceforth abbreviated to " French ", used a corpus of 1,900 telephone conversations, comprising 80,000 words altogether. These 80,000 words represent 2,240 *different* words, 737 of which occur in at least 1% of the conversations and together make up 96% of the total occurrences of words.

Different dialects are represented in these telephone conversations. It has not been possible to take this into account: the present writer transcribed the 737 words, which French *et al.* give in normal orthography, according to a stressed, dictionary pronunciation, as given by Kenyon and Knott. Where Kenyon and Knott list more than one form, the first listing was chosen, which is presumably the most common one. Examples are:

coffee, long: /ɔ/ only, not /ɑ/
what, want: /ɑ/ only, not /ɔ/
with: /ð/ only, not /θ/
always: /ɪ/ only, not /eɪ/, in the last syllable

The contrast between /ɔr/ and /oʊr/ was preserved, in accordance with the first variant given by Kenyon and Knott.

Only word-initial and word-final sequences are listed: no attempt was made to divide the disyllabic and polysyllabic words into syllables. Out of the 737 words, 371 are monosyllables. The average length of the words is, according to the authors, 1·23 syllables.

Since no inflected forms are included in the word list, the total number of syllable-finals listed here is probably considerably lower than the number which actually occurred in the telephone conversations: a number of final clusters ending in /t, d, s, z/ will be missing.

Malone deals with monosyllabic words only. Therefore, certain sequences are omitted ; for example, the initial sequence /splɛ-/ of *splendid* is missing.

His corpus is American English in general, and he includes material from different dialects. Thus, he lists both /æ/ and /ɑ/ for *ask,* and both /ɑ/ and /ɔ/ for *off*. All this material has been included in the present survey, except distributions which apply to " r-less " dialects only, such as the occurrence of /ɔ/ before the codas /dʒ, ɡ, θ, vz, mθ/ (*George, morgue, forth, wharves, warmth*).

Malone gives a supposedly complete list of all initial and final margins, but he does not state explicitly which particular combinations of syllable nuclei and margins occur.

He generally confines himself to statements about restrictions on their distribution. For the present study, the principle was adopted to include those combinations which are not explicitly ruled out. However, it seems likely that Malone's negative rules do not cover all the restrictions on actual occurrence. Thus, before complex codas beginning with /r/, except /rd, rz/, only six syllable nuclei are explicitly ruled out, viz. /i, ɪ, ɛ, æ, u, ə/ ; to these must be added /ɚ/, which Malone writes /ɜr/ ; but it is doubtful that all his other 7 nuclei occur there. Likewise, there seem to be no restrictions on the occurrence of /ʊlC(CC)/, but so far only /ʊld, ʊlf, ʊlft, ʊlfs, ʊlvz, ʊlz/ have been documented. Other doubtful cases are particularly the codas /pts, pθ(s), pst, kts, kst(s), ksθ(s), dθ(s), dzd, fts, fθ(s), mpst, mf, nzd, lfθ(s), lst, lʃ(t)/: it does not seem probable that they can combine with all the syllable nuclei which are not explicitly ruled out by Malone's negative rules. The same argument applies, to a lesser extent, to a number of other codas, such as /b, bd, bz, gd, zd, ndz, nθ(s), nst, lts, ltʃ(t), lkt, ldʒ(d), lf, lft, lfs, lθ, lm, lmd, lmz, ln, lnd, lnz/.

A few commonly occurring phoneme sequences may be left out by mistake, such as the codas /mfs, ndʒ, ndʒd, ŋd, rmθs/ and the nucleus + coda combinations /əŋ(CC), ɔld, ɚldz, oʊlz, ɪrs/.

O'Connor and Trim, henceforth abbreviated to " O'Connor ", based their study on the monosyllabic words of the Received Pronunciation (RP) of Southern England, with the exclusion of the following types of words:

(a) proper names
(b) learned and scientific terms
(c) anglicized foreign words
(d) rare and archaic words
(e) slang and interjections
(f) unusual pronunciations.

The authors point out that the decision as to which particular words should be excluded under headings (b) - (f) is necessarily arbitrary.

In part the particular phoneme sequences which occur are listed ; in part only rules are given for which phoneme sequences are permitted. There is no information about which particular syllable nuclei occur after each onset ; only their number is given. There is no information about the number of nuclei combining with each coda.

Trnka's corpus is *An English Pronouncing Dictionary* by Daniel Jones, and thus represents Southern British Received Pronunciation (RP).

Like Malone, he generally gives the restrictions on the distribution of phonemes, rather than stating which combinations actually do occur. However, Trnka's rules seem to cover the material more completely than Malone's. All combinations which are not explicitly ruled out are included in the present study. An exception is made for certain consonant combinations, at least 13, which appear to be permitted according to the rules, but which are not listed anywhere as actually occurring in combination with syllable nuclei, initially, medially, or finally, viz. /tʃj, tn, fn, θn, ʃm, ʃn, zn, mt, mʃ, mn, nl, nr, nw/. These are not included in the present survey.

There are certain restrictions on his corpus which are responsible for the relatively

low number of phoneme combinations in certain positions. (a) He takes into account monomorphemic words only, and a number of final consonant clusters, especially in /-t, -d, -s, -z/, are therefore absent ; these would be common inflectional endings. (b) Since his data are the words of a so-called r-less dialect, there are no sequences /-VrC(CCC)/. (c) He is concerned with the occurrence of syllable margins before and after *stressed* nuclei only.

On the other hand, in certain respects the type of his corpus results in a larger number of phoneme combinations than other studies. (a) Because his corpus is a dictionary, he includes a number of unusual, bookish pronunciations, especially for onsets, such as /tm/ in *tmesis*. (b) He includes morpheme-medial phoneme sequences, and the components of some of these may belong to different syllables. He is not concerned with determining the point of syllable division in words of more than one syllable, and with the occurrence of phonemes relative to the syllable ; he states restrictions on the sequential occurrence of phonemes only.

The present survey includes among initials and finals those of Trnka's consonants and consonant sequences which are initial and final relative to the monomorphemic word. In addition, Trnka lists 78 consonant clusters which occur morpheme-medially only. It is assumed that there is a syllable boundary somewhere in these sequences, and it is possible that if they were split up into syllable-final + syllable-initial sequences, a few additional initial and final margins would have to be added to the list.

On the other hand, as far as the combinations between syllable nuclei and margins are concerned, it is not possible to separate in his data word-initial from word-medial (CC)CV occurrence, and word-final from word-medial VC(CCC) occurrence. Since in non-initial (CC)CV and in non-final VC(CCC) the syllable boundary may well be in the middle of the consonant cluster, it is likely that there are fewer combinations than those listed when we consider position initially and finally in the *syllable* only. Thus, he lists /Vr/ and /ʒV/ combinations. /Vr/ does not occur utterance-finally in this dialect, and /ʒV/ is at least rare utterance-initially.

Wallace's corpus is 10,000 words of running conversation in a Mid-Western university community, tape-recorded, transcribed phonemically, and divided into syllables according to explicit rules.

There is some ambiguity as to the meaning of " initial " and " final " clusters, whether they comprise only word-initial and word-final clusters or whether they also include close-knit consonant combinations resulting from syllable division in word-medial consonant sequences. It is assumed that the latter is the case.

Wallace discusses only consonant clusters, not single consonants, and there is no information about combinations between syllable nuclei and margins.

Whorf sets up a " structural formula of the monosyllabic word in English (standard Mid-Western American) ". The present survey lists all the onsets and codas which are possible according to his formula.

Some well-established syllable margins are not accounted for by this formula, viz. the onsets /spj, skj/ and the codas /pts, kts, ksts, fts, nt, nts, ŋk, ŋkt, ŋks, lt, lts, lk, lkt, lks, lfθ, lfθs, lv, lvd, lvz, rnt/.

On the other hand, his formula seems to allow for far more codas than actually occur in English. This is mainly due to two facts: (1) /r/ is permitted before any otherwise occurring coda, without any restrictions ; (2) /t/ or /d/, /s/ or /z/, and /st/ or /zd/ are permitted after any otherwise occurring coda, the distribution of the two members of each pair being conditioned by the voicelessness/voicing of the preceding consonant. It appears to be in conflict with the structure of English that /r/ should be found before /ʒ, ŋ/ at all, and combinations like /rsp, rsk, rntʃ, rndʒ, rlp, rltʃ, rlb, rlf, rlθ, rls, rlm/ are also unlikely. As far as /t/ or /d/, /s/ or /z/, and /st/ or /zd/ are concerned, though these may be inflectional endings, it does not seem possible to add them to all the codas permitted by the formula ; /zd/ probably never occurs at all as such an ending and the 2nd person singular verbal suffix /st/ (" thou triumphst ") can hardly be considered an active suffix any more.

Whorf's formula is a model for generating monosyllabic words in English, whether they exist in the present vocabulary or not. For example, a new trade name might be coined on the basis of permitted sequences. However, some of the combinations which result from the formula appear to be in conflict with the phonological structure of English. A few were mentioned in the preceding paragraph. In addition, the codas /(r)sd(CC), (r)sg(CC), (r)mg(CC), (r)np(CC), (r)ng(CC), (r)lg(CC)/, which result from Whorf's " term 12 ", hardly occur in English, and /s/ or /z/, /st/ or /zd/ of term 14 cannot occur after /ʃ, ʒ/.

Of the 388 codas which are possible according to Whorf's formula, 214 are not mentioned by other analysts, and no examples can be found for them. Included among the 214 items are codas which do occur in idiolects where the 2nd person singular present verbal ending /-st/ is used, e.g. /rkst/ as in *barkst*.[1]

2.2. *Phonemic Normalization*

The data were normalized in phonemic interpretation and content. For the purpose of the present study, the English phonological system is considered to consist of the following segmental units.

There are 22 consonant phonemes, /p, t, k, b, d, g, f, θ, s, ʃ, v, ð, z, ʒ, m, n, ŋ, l, r, h, j, w/. [tʃ] and [dʒ], as in *chip* and *Jim*, are considered sequences of phonemes, /tʃ/ and /dʒ/, and [hw] or [ʍ], as in *white*, is interpreted as /hw/.

There are 15 syllable nuclei, /i, ɪ, ɛ, æ, ɑ, ɔ, ʊ, u, ə, ɚ, eɪ, aɪ, aʊ, ɔɪ, oʊ/, as in *beat, bit, bet, bat, bomb, bought, bush, boot, but, Bert, bait, bite, bout, boy, boat*. It is irrelevant for the purposes of this study whether some of these nuclei should be considered phonemically complex. If, for example, *beat, boot, bait, bite, bout* are interpreted as /bijt, buwt, bejt, bajt, bawt/ respectively (and *bit, bet, pot, put* are considered /bit, bet, pat, put/) /j/ and /w/ are part of the syllable nucleus, and do not belong to the syllable margin (coda). There will therefore be no final margins */j(CCC), w(CCC)/. The set of symbols which were chosen for the syllable nuclei should not be interpreted as suggesting any specific phonemic analysis. They are merely the most commonly accepted *phonetic* symbols for these units, in accordance with the rules of the International Phonetic Association.

[1] *For a list of these 214 codas, see Appendix A III in* SRL Report No. 5.

The dialect which forms the basis of this study is the relatively uniform type of English commonly called General American (GA). This dialect may be an abstraction, but such an abstraction is useful and necessary for anyone engaged in learning to speak English. A speech synthesizer has to be taught a similar " standard " dialect.

Data taken from other dialects were to some extent translated into GA. For example, the syllable nuclei [ɜ] and [ə] of *bird* and *father*, as they occur in the Received Pronunciation (RP) of Southern England, were replaced by /ɚ/ ; [ɜ] (and [ə] in some of its occurrences) and [ɚ] are manifestations of phonemes occupying similar positions in different dialect systems ; they do not contrast in any one dialect, RP or GA. Also, [ɑ] was substituted for RP and New England [ɒ] of *pot*, though there is a contrast between [ɑ] of *part* and [ɒ] of *pot* in such dialects, since this opposition is paralleled by the contrast [ɑr] *vs.* [ɑ] in GA ; /ɑ/ thus occurs in *bomb* /bɑm/, *balm* /bɑm/, *father* /fɑðɚ/, *farther* /fɑrðɚ/.

However, in all other cases, data from different dialects were used whenever it is a question of a difference in the distribution of well established phonemes. Thus, the sequences /ir/ and /ɪr/ ; /ɛr/, /ær/, and /eɪr/ ; /ɔr/ and /oʊr/ ; /ʊr/ and /ur/ were all included in the overall data, since they are listed by one or more of the sources for the present study ; however, they are not in contrast in one common variant of GA. Also, both /ɑg/ and /ɔg/, as well as /ɑŋ/ and /ɔŋ/ are included, since Kenyon and Knott list both possibilities for words like *dog, long*. Likewise, both /æ/ and /ɑ/ are noted for such words as *half, path, grass, dance*, and both /tju-/, /sju-/ and /tu-/, /su-/ are listed for *tune, suit*, etc.

It should be noted that only those dialects are considered which are represented in Kenyon and Knott's and in Jones' pronouncing dictionaries. There are doubtless other English dialects which cannot be accommodated within a 15-nucleus system, and where additional combinations between syllable nuclei and margins occur. Thus, Hill (1958) lists the initial combinations /kjɑ-/ and /gjɚ-/ for words like *car, girl* in Tidewater Virginia.[2] Such combinations are not included in the present survey.

In one respect important material from the above mentioned pronouncing dictionaries is omitted, *viz.* the lack of /r/ before consonants and junctures in the so-called r-less dialects ; no " centring diphthongs " are taken into account, for example, in *clear, chair, pour, poor*, and the codas in words like *far, court, large* are considered to start with /r/. However, only few additional nucleus + margin combinations would have to be added to take care of the r-less dialects. A cursory examination of the data yielded only five, /ɔdʒ, ɔdʒd, ɔmd, ɔmθ, ɔmθs/, as in *forge, forged, formed, warmth, warmths*. On the other hand, for r-less dialects one could cut down the number of codas and of nucleus + margin combinations drastically, since there are no /rC(CC)/ sequences.

This apparent juggling with the data was necessary, in an attempt to make this study as general as possible, while ensuring that the data from various sources could easily be compared.

In the " suggested minimum " (2.32) not all the variants mentioned above are

[2] *Re-interpreted according to the present phonemic analysis ; Hill's transcription is* /kyahr/, /gyəhrl/.

included. Only *one* dialect is taken into consideration, *viz.* a type of GA with a relatively simple phonemic system. This dialect has the 15 syllable nuclei and the 22 consonants listed above, and the following distributional characteristics should be noted.

Half, bath, grass, dance, etc., have the nucleus /æ/.

Can, auxiliary verb and noun, both have the nucleus /æ/ and are homonyms.

/ɑ/ is the nucleus of *balm* as well as of *bomb,* the two words being homonyms.

Dog, long, etc., have the nucleus /ɔ/.

There is no contrast between [ɔr] and [oʊr]; the phonetic sequence which occurs can be interpreted as either /ɔr/ or /oʊr/; arbitrarily, the transcription /ɔr/ was chosen.

The same argument applies to [ir] *vs.* [ɪr], [ɛr] *vs.* [ær] *vs.* [eɪr], [ur] *vs.* [ʊr]. The transcriptions /ɪr, ɛr, ʊr/ were chosen.

There are no onsets */tj, dj, θj, stj, sj, zj, nj,, lj/.

/r/ occurs before consonants and junctures, as well as before vowels.

Data from the various sources were normalized phonemically as follows.

Dewey: The two syllable nuclei in *above* (Dewey: [əbʊv]) are considered manifestations of the same phoneme, /ə/. His [ʊr] and [ər], as in *turn* and *utter,* are both interpreted as /ɚ/. Dewey considers the vowels of words like *alms, part, ma,* on the one hand, and of words such as *odd, not,* on the other, different phonemes. They are both interpreted as /ɑ/ in the present survey. Dewey's diphthong [iu], as in *few,* is interpreted as /j/ + /u/.

Malone: Malone distinguishes between /ɑ/, as in /ɑrm/ *arm,* /ɑsk/ *ask,* and /ɒ/, as in /ɒn/ *on.* Both are here considered instances of /ɑ/. [ɔɪ] is not considered a diphthong by Malone, and there are no data available on its combinatory possibilities. Thus, only 14 syllable nuclei are taken into consideration for his material. Malone's /ɜr/ is interpreted as /ɚ/. It is possible that his codas /rθt, rsts, rðz, rzd, rldz/ occur only after his /ɜ/, in words like *berthed, firsts, berths* (one pronunciation), *furzed, worlds,* and that they therefore do not occur at all according to the present analysis.

O'Connor: O'Connor and Trim's material cannot be normalized to the same extent. The authors postulate 24 consonant phonemes, considering /tʃ/ and /dʒ/ unit phonemes. They have 21 syllable nuclei, *viz.,* in addition to the 15 nuclei listed above, the centring diphthongs /ɪə, ɛə, ɔə, ʊə/, /ɒ/ contrasting with /ɑ/, and unstressed /ə/ as separate from stressed /ʌ/. One can therefore expect more combinations of nuclei and margins than for the type of GA which serves as basis for this study.

Trnka: Both his /ɑ:/ as in /fɑ:/ *far,* and /ɒ/, as in /stɒp/ *stop,* are replaced by /ɑ/, and his /ɜ/ of /bɜd/ *bird* is rendered by /ɚ/. His nuclei /ɪə, ɛə, ʊə, aɪə, aʊə/, as in /hɪə/ *here,* /ðɛə/ *there,* /pʊə/ *poor,* /faɪə/ *fire,* /paʊə/ *power,* are disregarded. No data are available on the combinatory possibilities of [ɔɪ], since Trnka considers this a sequence of the vowel /ɔ/ and the consonant /j/, and only 14 syllable nuclei are therefore considered for his material. None of the /j(CC)/ combinations which he lists are included in the present survey: they are all instances of /ɔɪ(CC)/ and /ʊɪ(CC)/ or /uɪ(CC)/, which Trnka interprets as /ɔj(CC)/ and /ʊj(CC)/, as in *hoist, ruin.*

Wallace : Her single phonemes /č, ǰ/ are interpreted as clusters, /tʃ/ and /dʒ/, and her /ərC(C)/ is analyzed as /ɚC(C)/.

Whorf : Whorf interprets /tʃ, dʒ/ as single phonemes ; they are analyzed as clusters in this study. /y, w/, which, according to his formula, may occur immediately after the vowel, are in the present interpretation part of the syllable nucleus, as in *buy, how*. His formula is re-interpreted accordingly.

2.3. *Overall Data*

2.31. *Maximum.* Data taken from various sources were added up in a " maximum " figure, giving the totality of phoneme combinations occurring in some type of the GA and RP dialects of English, restricted by the phonemic normalizations of 2.2.

A few additional restrictions were imposed in setting up the maximum list of phoneme combinations. Insufficient data are available on word-medial phoneme sequences, and the maximum list was therefore restricted to sequences initially and finally in the word. No phoneme sequence was included for which no example could be found. Thus, a number of word-final consonant combinations inferred from Whorf's formula, and some of the nucleus + margin combinations which are possible according to Malone's distributional statements, are excluded. Omitted are also certain sequences listed by Trnka which, it is suspected, occur word-medially only (see 2.1).

On the other hand, the maximum includes phoneme sequences which occur but rarely in GA and RP, such as some of the onsets quoted by Trnka.

This writer has not herself systematically excerpted any dictionary or any long corpus of running text ; it is likely that further research will increase the total inventory of phoneme sequences, but it does not seem probable that the figures will be radically different.

2.32. *Suggested Minimum.* In the maximum list of phoneme sequences, the frequency or probability of occurrence is disregarded. It is obvious, however, that they are not all of equal importance in the structure of English. This is clearly brought out by the fact that relatively few combinations are listed by Dewey, French, and Wallace, whose studies are based on limited corpora or on frequency of occurrence. One could generate perfectly intelligible English sentences from a model which would take into account far fewer permissible phoneme sequences. Thus, a device for speech synthesis using segments of some type as building blocks would not need an inventory of the size suggested by the maximum figures. We could reduce the figures considerably in the following ways.

(a) One need not take more than one dialect into account. Thus, since a common type of GA, e.g. in the Mid-West, has no initial sequences /tj, dj, θj, stj, sj, zj, nj, lj/, they may be excluded from our lists. Since the same dialect does not distinguish between [ir] and [ɪr] ; [ɛr], [ær], and [eɪr] ; [ɔr] and [oʊr] ; [ʊr] and [ur], followed by one or more consonants or by juncture, only one sequence of each set need be included in the inventory.

(b) Uncommon proper names, especially foreign names, rare words, and unusual pronunciations could be disregarded and omitted from our corpus. This is what

O'Connor and Trim have done in setting up the two-member initial and final phoneme sequences in English (2.3). One could go even further in restricting the vocabulary, and retain only those sequences which are necessary to account for the more common words such as we find them in word frequency lists.

A word of caution might be added about the adequacy of the sample size in frequency studies. Wang and Crawford (1960), in their evaluation of "Frequency studies of English consonants", found close agreement among such studies as those excerpted here for single consonant phonemes. A sample size of a few thousand units was sufficient to indicate first-order probabilities. For the probabilities of longer sequences, as they are involved here, a much larger sample would be needed. It is significant that, as far as consonant clusters are concerned, there is far from agreement among studies which are based on frequency counts or on a limited corpus (Dewey I, French, Wallace): their lists of syllable margins differ markedly, not only in the number of margins, but also in the particular items which are included.[3]

(c) In normal conversational style a great number of assimilations, elisions, and other types of simplifications take place, such as we see listed in phonetic text books, and as they are listed by Wallace. Thus, instead of /-pθs/ and /-ksts/, /-ps/ and /-ks/ are frequently used in *depths* and *texts*. We could take account of this fact in setting up our segment inventory and use such simplified sequences wherever possible.

This survey suggests a minimum number of sequences required for English. The minimum comprises the particular sequences which (1) are observed to occur in a common type of GA, (2) are needed in order to make oneself well understood in English, and (3) would be required in an inventory of segments for speech synthesis. The figures are conservative estimates. It is likely that more phoneme sequences could be excluded, giving an even lower minimum.

2.33. *Prosodies.* In computing the number of segments required for each type of synthesis, it is necessary to take into account varying conditions of prosody.

There is a complex relationship among the various prosodic features of English. Three types of prosodies are generally isolated, *viz.* length, stress, and intonation. Experimental research suggests (1) that it is difficult to define each of these prosodies in terms of physical or physiological parameters, and (2) that there is a high degree of interdependence between them. Some of the differences in the phonemic analysis of English prosodies may be due to this interdependence; different linguists divide the complex of prosodic features in various ways (Sivertsen, 1960).

2.331. Perception of *length* in speech sounds probably has a high correlation with physical duration. It is assumed that duration is not an independent variable in English, and that length differences may be predicted in terms of the segmental phones, stress, and intonation.

2.332. As many as four contrastive *stresses* have been postulated for English (Gleason, 1955; Trager and Smith, 1951). There is some doubt whether one has to postulate so many stresses in order to account for the data (Sivertsen, 1960). Some of the stress differences seem to be associated with intonational differences. There is

[3] *See Appendix A of* SRL Report No. 5.

no general agreement among linguists and phoneticians how to define stress. As far as acoustical parameters are concerned, stress seems to have some correlation with duration, intensity, and fundamental frequency, possibly also with other factors. Recent work on English prosodies, e.g. by Bolinger (1958) and Fry (1958), suggests that fundamental frequency may be more important than duration and intensity, not only for the recognition of the intended intonation pattern, but also for stress perception. Peterson (1958) reported a case where a girl was able to speak with apparently normal " stress patterns " in an iron lung, which did not permit her to control respiratory effort.

It seems likely that for speech synthesis it is necessary to simplify the prosodic system. As a working hypothesis it is therefore proposed that pitch differences only be incorporated in the segment inventory. It may be noted that Liberman *et al.* (1959) have adopted a different kind of simplification for their synthesizer: only stress contrasts are included, while intonational differences are disregarded. Only synthesis of extensive material and listening tests could decide whether the omission of stress, or of intonation, from the phonological system of the synthesizer language would result in any serious loss of intelligibility.

The present survey thus takes no account of stress differences other than those subsumed in the intonation system. The alternative might be to double the inventory of phoneme sequences, at least all those which contain syllable nuclei, allowing for one stressed and one unstressed variant of each segment.

2.333. It seems necessary to take account of *pitch* differences in speech synthesis, not only because it seems to be an important part of " stress " contrasts. Monotone speech, especially if it is synthesized speech, does not only sound unnatural, but is also less intelligible than speech with normal intonation. The poorer intelligibility might not appear in articulation scores on lists of monosyllables, but it would show up in whole utterances and connected discourse. This is so because the relationships between the various parts of the utterance, the attitude of the speaker to what he is saying and to his interlocutor, and various kinds of " unsaid " things are conveyed by means of intonation. The meaning conveyed by the intonation is not an unimportant addition to the meaning of the utterance, but an essential, integral part of that meaning.

The intonation of American English has been analyzed as a number of patterns consisting of combinations of four pitch levels (Pike, 1945), and, more recently, three terminal contours or junctures (Gleason, 1955 ; Hockett, 1955 and 1958 ; Trager and Smith, 1951). This type of analysis is accepted for the purposes of the present study.

The number of intonation patterns is large. Pike lists, for the intonation " centre " or " nucleus ", four level contours, six falling contours, six rising contours, nine falling-rising contours, and one rising-falling contour, i.e. a total of 26 contours, and each of these can be combined with various pitch levels in the remaining part of the intonation pattern.

All of these intonation patterns are probably not equally important. It is likely that one can be understood even if one uses only a limited number of them. Thus, the English Language Institute of the University of Michigan, in its intensive courses for

foreign students, uses three pitch levels and four pitch glides only, and no attention is paid to the " terminal contours " as distinct from combinations of pitch levels. The pitch levels are 1, 2, and 3, numbered from the lowest pitch, and the gliding contours are 3-1, 3-2, 2-3, and 3-1-2.

It does not seem likely, at the present stage, that one could cover in speech synthesis the whole range of intonational contrasts in English. In so far as intonation is to be simulated at all, it seems better to base the synthesis on a simplified intonation model, and the particular system adopted by the English Language Institute might serve as this model.

There are great problems in applying even a simplified intonation model to speech synthesis, however.

First, the pitch levels which are postulated do not represent absolute pitch values. They are relative values only, and syllables which are analyzed as having the same phonemic pitch level may in fact have widely different absolute pitch values. There is, in fact, extensive overlapping between the phonemic pitch levels even in the speech of a single speaker speaking on one occasion to one audience in one style.

Second, instrumental analysis shows that actual speech is not a succession of pitch levels, but seems to be continuously changing in pitch. The postulated pitch levels are abstractions from this continuum.

The latter difficulty could probably be taken care of when the segments to be used are from recorded natural utterances, since these must be assumed to include the pitch dynamics. If the segments are to be produced electronically, the pitch fluctuations will have to be simulated, after a careful study of actual speech. In either case, one would have to match the pitch carefully where two segments abut. This is done more easily electronically than with segments from recorded natural utterances, since one would need utterances where one cut yields both the best formant match and the best fundamental frequency match.

It does not seem possible to simulate, in synthesis based on segments, the great variety of actual pitch values. One will probably have to accept medium or standard pitch values for each phonemic pitch level. This does not mean that all segments in our inventory which have the same phonemic pitch level will necessarily have the same fundamental frequency, however. The data presented by Peterson and Barney (1952) show that certain (especially high) vowels will naturally have a higher fundamental frequency than other (low) vowels. Consonants may also have different frequencies. For one study of speech synthesis from phoneme dyads, taken from recorded natural utterances, at the Speech Research Laboratory of the University of Michigan, the following frequency values were chosen for a particular male speaker.

Pitch level /1/: the syllable nuclei vary from 101 to 115 cps., each having its own specific value ; the consonants have a value of 115 cps.

Pitch level /2/: nuclei 124-141 cps.; consonants 132 cps.

Pitch level /3/: nuclei 168-190 cps.; consonants 168 cps.

Even with these adjustments one can probably not avoid a certain sing-song effect when standardized pitch values are used. Whether this will affect intelligibility remains to be seen.

3. Syllable nuclei and margins

3.0. When one applies the immediate constituent analysis to the syllable, it is found to consist not of a string of phonemes, but of a sequence of higher-layered units. The immediate constituents of the syllable are the nucleus (N) and the margin (M). In English every syllable has a nucleus, and in addition the syllable may contain an initial margin, called an onset when it follows a juncture, and/or a final margin, called a coda when it precedes a juncture. Nucleus and margins may be simple or complex (Hockett, 1955 and 1958).

It seems reasonable to assume that these immediate constituents of a syllable could be used as basic units for speech synthesis. A segment inventory for such a synthesizer would include all the common syllable nuclei and margins occurring in that particular language.

For a common type of GA one would need 15 nuclei, /i, ɪ, ɛ, æ, ɑ, ɔ, ʊ, u, ə, ɚ, eɪ, aɪ, aʊ, ɔɪ, oʊ/. The number of syllable margins is much larger, and will be considered in the following pages.

3.1. *Initial and Final Margins*

Table 2 is a survey of the number of syllable margins found in each of the publications consulted. The maximum (2.31) and suggested minimum (2.32) are also given.

Table 2

	INITIAL	FINAL
Dewey I	65	88
Dewey II	71	182
French	40	49
Malone	75	186
O'Connor	71	140
Trnka	85	66
Wallace	55	94
Whorf	65	388
Maximum	118	238
Suggested minimum	66	168

Number of syllable margins in English, according to various sources.

The margins listed are either initial or final in the syllable. Some margins occur both in initial and final position. Consonant sequences which occur only utterance-medially between syllable nuclei are not included. " Initial margin " and " final margin " are not synonymous with " onset " and " coda ". An onset is a syllable-beginning occurring after juncture, and a coda is a syllable-ending occurring before juncture. Since both Dewey and Wallace include syllable-beginnings and -endings

which are found word-medially, it is likely that some of their consonant combinations are part of interludes (Hockett, 1955 and 1958), i.e. of intervocalic consonant sequences not interrupted by juncture, and therefore by definition not onsets or codas.

In an evaluation of the figures of Table 2, the following points should be kept in mind.

Dewey I: Dewey's figures are 51 and 96. For the initials, I have added zero onset and /pj, tj, kj, bj, dj, gj, fj, sj, vj, mj, nj, lj, hj/, all of which occur, but which he interprets as C + the first half of a diphthong [iʊ]. For the finals, I have added zero coda, and deducted /rtʃ, rks, rv, rmz, rnd, rl, rld, rldz, rlz/ ; these are examples of ɚ C(CC) only.

Dewey II: Dewey's figures are 57 and 181. I have added the same 14 onsets as for Dewey I. His data do not enable us to decide whether there were more /(C)Cj/ onsets, such as /θj, stj, skj, zj/. For the finals, I have added zero coda.

O'Connor: The figure for the finals has been computed from his rules on pp. 118-9, which give 1 zero final, 20 C, 62 CC, 50 CCC, and 7CCCC. On p. 114 only 59 final CC combinations are mentioned, and these presumably include the last part of (C)CCC finals. It is difficult to explain this inconsistency.

Wallace: Wallace's figures are 32 and 37, but apply to consonant *clusters* only, exluding /tʃ, dʒ/. It is assumed that the following margins also occur in her data: zero onset ; 20 simple onsets, /p, t, k, b, d, g, f, θ, s, ʃ, v, ð, z, m, n, l, r, h, j, w/ ; the complex onsets /tʃ, dʒ/ ; zero coda ; 18 simple codas, /p, t, k, b, d, g, f, θ s, ʃ, v, ð, z, m, n, ŋ, l, r/ ; the complex codas /tʃ, dʒ/. This gives a total of 55 onsets and 94 codas. It is possible that some of the /rC(CC)/ finals, such as /rtʃ, rg, rnt, rnd, rl, rlz/, should have been deducted, since it is possible that in her limited corpus they would be examples of [ɚ C(CC)] only (cf. 2.2).

Whorf: the coda figure is probably too high: it includes 214 codas which are not mentioned by other writers (cf. 2.1).[4]

Since Dewey I and French are based on frequency counts, and include only the most frequently occurring margins, their figures could be expected to be relatively low. This applies particularly to French, where the figures obtain for position initially and finally in the word only, not in the syllable. Wallace's figures might be expected to be lower than Malone's, O'Connor's, Trnka's and Whorf's, since her corpus is more limited than theirs. The small number of syllable-finals found in Trnka's material is explainable from the fact that only monomorphemic words are included in his survey, so that the clusters resulting from the addition of inflectional endings, especially /-t, -d, -s, -z/, are excluded. Also, this is an " r-less " dialect, with no /r(CCC)/ finals.

In the " maximum " figures are included all the margins of which examples are found initially and finally in *words*. Excluded are some finals mentioned by Malone which, it is suspected, do not occur when [ɚ] is interpreted as / ɚ / and not as /ər/ (2.1, 2.2). Excluded are also 214 codas which seem possible according to Whorf's formula, but which are listed by no other writer (2.1). On the other hand, the maximum includes margins which do not occur in the types of English with which the present writer is

[4] *These 214 codas are listed in Appendix A III of* SRL Report No. 5.

familiar, and which probably are exceedingly rare, such as the initials /ps, psj, pʃ, tm, kn, bd, ɡn, zd, mw/ in words like *psyche, pseudonym, pshaw, tmesis, knout, bdellium, gnosis, 'sdeath, moiré,* all quoted by Trnka, or the initials /pt, pf, zl, zw/ as in *pterosaur, pfennig, zloty, zouave,* which we find in Jones' pronouncing dictionary, or other initials mentioned by various writers, such as /pw, bw, fθ, sfr, sθ, snj, ʃn, zbl, mn, nw/ as in *pueblo, bueno, phthisis, sphragistic, sthenia, Snewin, schnapps, 'sblood, mnemonic, noire.* Other initials occur in foreign names only, and are likely to be pronounced only by speakers with some knowledge of the language in question, such as the initials /dv, dn, vl, vr, ŋ, ŋg/ in names like *Dvorak, Dnieper, Vladivostock, Vryburg, Ngami,* all found in Jones' dictionary. Similar arguments apply to some of the final margins. It is probable that additional rare syllable margins may be found by a study of foreign names and of the vocabulary of specialist scholars. Thus, Truby (1959) suggests—and rejects—the initials /pnj/ and /ɡd/ for *pneumatic* and *Gdynia,* respectively.

The " suggested minimum " is the number of syllable margins considered necessary for the type of GA described in 2.2. It will account for the word-initial and word-final syllable margins found in the whole English vocabulary, with the exception of foreign names like those quoted above. To obtain this figure the " maximum " number was reduced in the following ways.

(a) Syllable margins which do not occur in the type of GA under consideration, such as the initials /tj, dj, θj, stj, sj, zj, nj, lj/, are excluded.

(b) Certain margins which occur exclusively in extremely rare words that are used only by a small minority of English speakers, if they ever do occur, are also disregarded. Examples are the initials /pʃ, pf, snj, zd/ in words like *pshaw, pfennig, Snewin, 'sdeath.* It is likely that most of these clusters will be replaced by other and more common margins with many speakers.

(c) Unusual margins are replaced by more common sequences which generally take their place, e.g. /ps/ > /s/ in *psyche,* /mn/ > /n/ in *mnemonic,* /pw/ > /pʊ/ or /pu/ in *pueblo.*

(d) Certain complex margins are replaced by other margins which frequently take their place, when the functional load of the contrast between them seems to be very low. Often the two clusters cannot be seen to contrast at all, but merely to be freely varying in all the morphemes in which they occur. Examples are:

/-dθ/	>	/-tθ/	as in *width*
/-fθs/	>	/-fs/	as in *fifths*
/-mpf/	>	/-mf/	as in *nymph*
/-mt/	>	/-mpt/	as in *tempt*
/-nʒ/	>	/-ndʒ/	as in *plunge*
/-ŋt/	>	/-ŋkt/	as in *linked*
/-rmpθ/	>	/-rmθ/	as in *warmth*

When there is a choice between two such clusters, it may be appropriate for other idiolects or dialects to choose the one which has been omitted here. For example,

/-ntʃ/ is chosen here for *lunch,* but /-nʃ/ may be a better choice in other dialects; both are not necessary in any one dialect.

These simplifications should not result in any loss of intelligibility, even in isolated monosyllabic utterances. It is likely that one could reduce the figures even more without any loss of information, especially in running texts where the context will contribute to an understanding of the utterance. A comparison between the suggested minimum, on the one hand, and the figures based on word frequency counts (Dewey I and French) or on everyday conversational material (Wallace), on the other, in Table 2, suggests that further reductions could be achieved by limiting the vocabulary of the synthesizer to more common words.

A complete list of all the syllable margins included in the "maximum" of Table 2, and their adoption, or reasons for their rejection, in the "suggested minimum" are given in the Appendix to this paper.[5]

3.2. *Medial Margins*

The figures quoted from French, Malone, O'Connor, Trnka, and Whorf apply to position initially and finally in the *word* only. There may be syllable-initials and syllable-finals which are found word-medially only. Dewey and Wallace include such intervocalic consonant sequences in their figures, but their corpus is limited and does not include all possibilities. Wallace does not seem to have found any utterance-medial consonant sequence which could not be divided into utterance-final and utterance-initial margins, though the case is not too clearly stated. In Dewey's material it is not possible to separate those syllable-initials and syllable-finals which occur in word-medial position from those which occur initially and finally in the word.

Trnka gives a list of all those consonant sequences which occur medially in the monomorphemic words listed in Jones' dictionary. There are 170 such sequences (plus, presumably, 22 single consonants), 78 of which occur neither initially nor finally. These sequences are intervocalic, consisting, it is assumed, of a syllable-final plus a syllable-initial consonant or consonant cluster. Since there is no indication of the point of syllable division, we do not know whether the data yield additional syllable-initials and syllable-finals.

The question should be asked whether our list of syllable-initials and syllable-finals will be able to account for intervocalic consonant sequences which must be considered interludes. Wallace discusses some of the problems involved in syllable division, and concludes that some sequences are ambiguous. The problem is also discussed by Hill (1958, pp. 84-8). It appears that there may be actual contrast between interludes and consonant sequences interrupted by juncture (Hockett, 1958, pp. 55-9; Lehiste, 1959; Sivertsen, 1960, pp. 21-3). However, in a list of those initials and finals which are necessary to produce reasonably intelligible speech, a list such as we might need in speech synthesis, we need not take these contrasts into account, since their functional

[5] *Appendix A of* SRL Report No. 5 *also lists the occurrence or non-occurrence of the syllable margins in the various sources, and includes those margins which are not contained in the maximum.*

load seems to be low. Thus, it is doubtful that pairs like *white shoes* and *why choose,* or triplets like *nitrate, night rate,* and *Nye trait* are usually distinguished in normal conversational style as distinct from pronunciation under laboratory conditions.

Likewise, though additional consonant sequences may be found in running conversational style, where assimilations are common and words are fused together, as in ['dountʃu] *don't you,* ['hæpm] *happen,* ['doump bə'liv] *don't believe,* ['kubm 'faın] *couldn't find,* such a style is not essential: there is not likely to be any loss in intelligibility if assimilations are avoided and the words are pronounced as separate units, with juncture between, even though such a style may sound slow, pedantic, and over-careful.

3.3. *Segment Inventory*

A segment inventory for speech synthesis based on syllable nuclei and margins would thus consist of the following segments: 15 syllable nuclei, 66 onsets, and 168 codas. The lists of onsets and codas share 23 items, /p, t, tʃ, k, b, d, dʒ, g, f, θ, s, sp, st, sk, ʃ, v, ð, z, m, n, l, r/ and zero. If one assumes that the same segments could be used as onsets and codas—which is by no means obvious, since consonants in initial and final position seem to have different characteristics (Lehiste, 1959)—one could reduce the total number of segments by 23. The margin zero would be the equivalent of certain types of phonemic juncture. One would be left with 15 nuclei and 211 margins, i.e., a total of 226 segments.

One would have to multiply this number by the number of prosodic conditions required for each segment. If one assumes that 3 pitch levels and 4 pitch glides will be adequate to account for the essential intonation contrasts in GA, and that no stress or length differences will be needed (cf. 2.33), the total number of segments could be computed as follows:

Voiceless phonemes and phoneme sequences need be represented by one segment only, irrespective of the pitch pattern involved. For all voiced phonemes or phoneme sequences we need at least 3 segments to account for the three pitch levels. For gliding contours, there is a complicating fact. In English, the distribution of pitch levels and glides is generally stated with reference to the syllable rather than to vowels and consonants, or to nuclei and margins. That is, the domain of a pitch glide is the whole syllable. It is reasonable to assume, however, that the main portion of the glide occurs on the nucleus of the syllable. Spectrograms show that the beginning and end of the glide may be distributed over the preceding and following margin, but the untested hypothesis is suggested that this may be an incidental phonetic feature of the utterance which is not essential for intelligibility. Thus, in the segment inventory for speech synthesis one could disregard pitch glides for syllable margins, and, for example, *dreams* with the intonation contour 3-1 might be synthesized in the following way: ^{3}dr + ^{3}i^{1} + ^{1}mz. The total number of segments would then be:

Margins involving only voiceless phonetic units	44	×	1	=	44
Margins involving at least one voiced phonetic unit	167	×	3	=	501
Nuclei	15	×	7	=	105
Total				=	650

Among the voiceless margins above are included the onsets /tr, hj, hw/, which are commonly phonetically voiceless. It is possible that some or all of the onsets /pl, pr, pj, tw, kl, kr, kj, kw, fl, fr, fj, θr, θw, spl, spr, spj, str, skr, skj, skw/ are also, at least with some speakers, for practical purposes completely voiceless, so that they could be included among the phoneme sequences which need only one prosodic condition. This would result in a reduction of 40 in the total segment inventory.

On the other hand, if the same segments cannot serve as onsets and codas, the 23 items omitted from our list would have to be added again and multiplied by the appropriate number of prosodic conditions. The total number of segments would then be:

Voiceless margins	56	×	1	=	56
Voiced margins	178	×	3	=	534
Nuclei	15	×	7	=	105
Total				=	695

It is likely that speech synthesized from segments consisting of syllable nuclei and margins would be more intelligible than synthesis from single phonemes or allophones (vowels and consonants), since the dynamics within the complex nucleus and the complex margin would be subsumed. However, one would still be faced with the problem how to make the transitions MN, as in [traɪ] *try*, NM, as in [ɑrm], *arm*, and MM in interludes, as in [-kstr-] *extra*, intelligible and natural. One could build smoothing circuits into the synthesizer ; whether this would make the synthesized speech intelligible remains to be seen. For the transitions NN, as in [aɪ ou] *I owe*, and MM across word boundary or potential juncture, as in [-tʃ s-] *speech synthesis* or [-ktr-] *electronic*, the abrupt change would probably do less harm. It would be heard as a juncture, and this effect could be reinforced by inserting a short segment of silence.

4. Half-syllables

One could conceivably synthesize speech from segments of half-syllable length. In order to establish the set of segments required, it is necessary to study the particular sequences of syllable nuclei (N) and margins (M) which occur in English.

4.1. *Initial MN and Final NM Sequences*

Table 3 is a survey of the number of initial MN and final NM combinations found in each of the publications consulted. Wallace and Whorf have no data on this point.

" Initial " and " final " are used in the same sense as in section 3. In principle, the sequences are initial and final relative to the syllable. Only in Trnka's data may the whole or part of the margin belong to another syllable.

The following points are of importance for an evaluation of the figures.

The figure quoted for O'Connor and Trim's study is a hypothetical one, and not directly deducible from their data. For initial MN combinations they specify how

TABLE 3

	INITIAL MN	FINAL NM
Dewey I	397	375
French	242	191
Malone	675	1443
O'Connor	892	
Trnka	692	477
Maximum	851	1241
Suggested minimum	713	938

Number of combinations of syllable nuclei (N) and margins (M) in English in relation to the syllable, according to various sources.

many syllable nuclei combine with zero onset (19) and with complex onsets (411), but there is no similar information for simple onsets. If one assumes that all of their 21 nuclei can occur with each of the 20 simple onsets /p, t, k, b, d, g, f, θ, s, ʃ, v, ð, z, m, n, l, r, h, j, w/ and with /tʃ, dʒ/, giving 462 additional MN combinations, there will be a total of 892. The figure is probably somewhat too high, since it is not likely that all CN combinations occur.

The " maximum " number of MN and NM sequences represents the combinations of which this writer has been able to find examples in word-initial and word-final position. Though additional combinations can probably be found, it does not seem likely that the figures will be radically different. Only Malone and O'Connor show higher figures, and as is suggested above (2.1 and 4.1), these figures are probably too high. Some of the figures making up the maximum are tentative only. Thus, the onsets /ʃp, ʃt, ʃk, ʃm, ʃn, ʃl/ are mentioned by Hill (1958) as occurring in proper names, but we do not know with how many nuclei they combine. Syllables containing syllabic consonants are disregarded, or rather they are interpreted as allophonic variants of /ə/ + consonant.

The " suggested minimum " is the number of MN and NM combinations considered necessary for the type of GA described in 2.2. The " maximum " number was reduced according to the same principles as for margins only (3.3).

A complete list of all the particular combinations of syllable nuclei and margins included in the " maximum " of Table 3 is given in the Appendix to this paper, with key-words for all word-initial MN and word-final NM combinations which have been documented in the GA and RP dialects of English.[6]

4.2. *Medial Sequences*

The minimum list does not account for all M and N combinations which occur word-medially, or rather, non-initially in the word in the case of MN, non-finally in the word in the case of NM. Thus, though /psN, pʃN, tsN, djN, gnN, ʒN, ŋN, ljN/

[6] *Appendix B of* SRL Report No. 5 *also lists the occurrence or non-occurrence of these combinations in the various sources, and includes those combinations which are not contained in the maximum.*

may be ruled out for word-initial position, we shall need these sequences for word-medial position, as in *trapesing, Hampshire, pizza, Indian, ignore, leisure, hanging, failure,*[7] and with our present data it is not possible to state how many different N's occur after each M.

Further, though some specific MN sequence may be non-occurring, or non-essential, word-initially, it may be needed word-medially, such as /blɔɪ, gjʊ, splɔ, splɔɪ, sploʊ, strɔɪ, vjɚ / in *tabloid, argument, explorer, exploit, explosion, destroy, behaviour.* Similar arguments apply to NM combinations. Trnka (1935, pp. 37-40) lists 238 NM sequences which occur only non-finally in the word (there are no similar specifications for those MN sequences which occur non-initially in the word), but M in many cases clearly spans two syllables. No other data on the subject seems to be available.

In the "maximum" and the "suggested minimum" there are included only combinations which have been attested or, in the case of the suggested minimum, which are considered essential initially and finally in the *word.*

4.3. *Segment Inventory*

This inventory of essential MN and NM sequences could be used in determining the segments required in speech synthesis using half-syllables as basic units. The segments would be joined together at the point of syllable division and in the middle of the syllable, i.e. between coda and onset and in the relatively sustained part of the syllable nucleus. Thus, the segments $M\frac{N}{2}$ and $\frac{N}{2}M$ would include the whole of M and half of N.

The advantage of such a system is that it would include the dynamics within complex M's and in the transition from M to N and from N to M. But several problems remain unsolved.

One problem would be to determine the best point for making a cut in the middle of N. In accordance with the basic principle underlying the dyad concept, we would like to make the cut, and join segments, where there is as little change as possible in the spectrum. In monophthongs this point is approximately in the middle of the phonetic unit, where the articulating organs and the formant frequencies have reached their target positions. Such syllable nuclei as /i, ɪ, æ, ɑ/ could thus probably be cut in the middle, and rejoined with other half-nuclei of the same kind without too much distortion, e.g. dri + imz = [drimz], where i in dri contains the first part of an [i], and i in imz contains the last part of an [i]. The dyad approach to synthesis assumes that this is possible.

It is more difficult to make a cut in the middle of a diphthong, or an essentially glided nucleus, however. If there is no sustained part in such a glide, one would be forced to make the cut in the middle of the glide. In a type of synthesis where the segments are produced and controlled electronically, this will cause less trouble than

[7] *The syllable division suggested here and in the following paragraph is arbitrary. It has no bearing on the present argument whether another syllable cut would reflect the phonetic facts better: in whichever way word-medial consonant sequences are divided, there will probably be some MN and NM combinations which do not occur word-initially and word-finally respectively.*

when the segments are to be cut out of recorded actual utterances. In the latter case it would be difficult, though probably not impossible, to match the spectra of the half-nuclei in $M \frac{N}{2} + \frac{N}{2} M$. However, for many dialects sustained parts can be found in the diphthongs. Lehiste and Peterson (1960), in spectrographic studies, found two target positions, one initial and one final, for [aɪ, aʊ, ɔɪ], and both were sustained, with a long glide between ; in [eɪ] there was an initial " inflectional point ", i.e. target position, but only the final part could be sustained, whereas it was only the initial part of [oʊ] which might be sustained ; the better part of both [eɪ] and [oʊ] was taken up by a gradual glide.

Wang and Peterson (1958), in setting up their dyad inventory, considered [aɪ, aʊ, ɔɪ] sequences of the syllabics [ɑ, ɔ] and the glides [j, w], and so were able to synthesize, for example, *buy* out of the segments $\frac{b}{2} \frac{a}{2} + \frac{a}{2}$ j. [eɪ] and [oʊ], on the other hand, were considered units which could be cut in the middle, without any discussion of the problems involved in cutting into the middle of a glide.

For half-syllable segments, the best solution appears to be to make a cut in one of the sustained parts of [aɪ, aʊ, ɔɪ], possibly in the first one, so that for [baɪt] one would need the segments b $\frac{aɪ}{2} + \frac{aɪ}{2}$ t, the first $\frac{aɪ}{2}$ representing one half-length of [a] as part of the diphthong [aɪ], the last $\frac{aɪ}{2}$ being the last part of [a] plus the glide to [ɪ] and the sustained part of [ɪ]. It does not seem advisable that the first part of [aɪ, aʊ, ɔɪ] should be replaced by the first part of [ɑ] and [ɔ], as in *pot* and *bought,* since the phonetic difference between them might be too great. In [eɪ] and [oʊ] the cut could be made in that part which happened to be sustained. Some English dialects or idiolects may presumably have no sustained part in the diphthongs, and the same situation will probably be found in other languages. Further research is required to study the results of segmenting into the middle of such a glide.

Another problem in synthesis from half-syllables is to determine the point where the two margins meet in $\frac{N}{2} M + M \frac{N}{2}$, where MM is an intervocalic utterance-medial consonant sequence. In most cases such a sequence can be divided into syllable-final + syllable-initial, or coda + onset, without loss of intelligibility (cf. 3.2). In complex interludes one could divide the consonant sequence at the point between two phonetic units which is closest to the syllable boundary, even though the syllable boundary in actual speech may be somewhat floating.[8] Most favoured points of syllable division may be established for English (Sivertsen, 1960, pp. 14-21). Thus, in [ˈɛkstrə] *extra*

[8] *The argument in the present section about the point of syllable division is not based on experimental data. It is assumed that, in English, phonetically trained listeners can determine the boundary between certain phonetic units generally called syllables, and that these boundaries are paralleled by certain physiological and acoustical facts.*

the syllable cut could possibly be made ['ɛks-trə]. It seems unlikely that this would result in lower intelligibility.

For simple interludes the problem is more complicated. The syllable boundary is neither before nor after, but in the middle of [t, b, l, r] in ['bɛt ɚ] *better*, ['rɑbɪn] *robin*, ['fɑlou] *follow*, ['vɛrɪ] *very*. After so-called long vowels and diphthongs and after unstressed vowels, the syllable division might possibly be made before the consonant, as in ['si-lɪŋ] *ceiling*, ['daɪ-vɪŋ] *diving*, [ə'raund] *around* (Sivertsen, 1960, pp. 14-21), but this could not be recommended for simple margins after stressed short vowels.

Two solutions might be suggested. (a) One could presumably double the medial consonant, and let, for example, *robin* be synthesized from $r\frac{\alpha}{2} + \frac{\alpha}{2}b + b\frac{I}{2} + \frac{I}{2}n$. This would probably not sound very natural, but it remains to be seen whether it would have low intelligibility. (b) A better solution might be to set up an extra set of segments for such cases, $\frac{M}{2}\frac{N}{2}$ and $\frac{N}{2}\frac{M}{2}$, including half-length phonetic units of M as well as of N. Thus, one would make a cut in the middle of the simple interlude, for example, in the middle of [t, b, l, r] in ['bɛt ɚ, 'rɑbɪn, 'fɑlou, 'vɛrɪ]. Even so, one could probably not use the flap allophone of /t/ for this position, but would have to accept a voiceless stop allophone, so that one could make the cut in the middle of the closure.

If we accept that $\frac{M}{2}\frac{N}{2}$ and $\frac{N}{2}\frac{M}{2}$, where M is a simple margin, i.e. a single consonant, have to be added to our segment inventory, the maximum number of extra segments will be $15 \times 22 \times 2 = 660$, i.e. the number of syllable nuclei multiplied by the number of consonants for both MN and NM sequences. However, such extra segments will be needed for the synthesis of -N_1CN_2- only when N_1 is stressed and one of the so-called short vowels, /ɪ, ɛ, æ, ɑ, ʊ, ə/, when N_2 is unstressed, and when C is not one of the consonants /h, j, w/. On the basis of the occurrence of CN in word-initial and NC in word-final position (cf. the Appendix), it is suggested that 239 $\frac{C}{2}\frac{N}{2}$ and 89 $\frac{N}{2}\frac{C}{2}$ segments should be required, i.e. a total of 328. To this figure should be added at least 6 $\frac{C}{2}\frac{N}{2}$, accounting for the sequences /ʒɪ, ʒə, ʒɚ, ŋɪ, ŋə, ŋɚ/, which do not occur in word-initial position, and an unknown, though probably not very large, number of segments to include NC sequences which do not occur word-finally, such as /ɪʒ, ɛʒ, æʒ/ in *vision, measure, casual.* For the purposes of this discussion, we shall assume that 245 $\frac{C}{2}\frac{N}{2}$ and 100 $\frac{N}{2}\frac{C}{2}$ segments are needed.

It is likely that the total inventory of M $\frac{N}{2}$ and $\frac{N}{2}$ M segments will have to be increased still further, in order to take into account those MN and NM sequences in

general which occur word-medially only, since these are not included in our lists. The present data do not allow us to estimate how great an increase this will mean. If all syllable margins could combine with all syllable nuclei, the total number of segments would be:

66 onsets	× 15 nuclei	=	990 segments
168 codas	× 15 nuclei	=	2,520 segments
Total		=	3,510 segments

Moreover, it is entirely possible that some or all of the syllable nuclei will have to be represented by two or more allophones, for example, one for stressed position and another for unstressed position. It is only by testing the segments in actual speech synthesis that one would get an answer to this problem. However, even if two allophones, one stressed and one unstressed, are required for each syllable nucleus, the total number of segments required will not have to be multiplied by as much as 2, since it is hardly likely that both allophones can combine with all syllable margins.

Disregarding the unknown number of extra segments which might be required according to the last paragraph, it is now possible for us to compute the number of half-syllable long segments required for speech synthesis:

$M \frac{N}{2}$	713
$\frac{N}{2} M$	938
Total	1,651

If we accept that $\frac{C}{2}\frac{N}{2}$ and $\frac{N}{2}\frac{C}{2}$ segments are needed, the total may rise to 1,996:

$M \frac{N}{2}$	713
$\frac{C}{2}\frac{N}{2}$	245
$\frac{N}{2} M$	938
$\frac{N}{2}\frac{C}{2}$	100
Total	1,996

This figure will have to be multiplied by the number of prosodic conditions required for each sequence. It is assumed that 7 intonation contours, i.e. 3 pitch levels and 4 pitch glides, will be adequate to account for all the essential prosodic contrasts in GA (cf. 2.33). We shall need all three levels for each of the half-syllable phonetic sequences. It is also likely that the four glides will be needed for each sequence. Though little work has been done on the distribution of pitch glides, or of intonation contours involving pitch change, over the various parts of the syllable, and over larger segments of the utterance, it seems hardly possible, for example, to crowd the whole pitch glide of a /3-1/ intonation contour into the first half of a syllable, and let

the last half have pitch level /1/. It appears that it will be necessary to segment in the middle of the pitch glide, so that the first part of it is included in the first half-syllable, and the last part in the last half-syllable. It is unfortunate that one cannot follow the same principle as with respect to the spectrum, and make the cut at the point of least change, since it may be difficult to construct segments which match in pitch in the middle of a pitch glide, particularly if segments cut out of recorded natural utterances are used. However, cutting at the point of minimum pitch change does not seem possible unless one is prepared to concentrate the whole pitch glide in one half of the syllable, as suggested above.

Assuming that all the intonation contours are required for each phonetic sequence, our segment inventory will comprise a total of 1,651 × 7 = 11,557, or 1,996 × 7 = 13,972. The latter number would be required if $\frac{C}{2}\,\frac{N}{2}$ and $\frac{N}{2}\,\frac{C}{2}$ segments are included.

5. Syllables

Another possible unit for speech synthesis based on segments would be the syllable. It is assumed here, as elsewhere in this study, that the syllable is a phonological and/ or phonetic unit in English, and that it can be isolated and defined.

There are various ways of estimating how many segments would be required. One could multiply the number of essential initial half-syllables with the number of final half-syllables, as estimated in 4.3:

$$713 \times 938 = 668{,}794$$

This figure would undoubtedly be far too high for speech synthesis segments. It is not likely that all these combinations occur at all in English, and they would certainly not be included among the more common syllables that would be required as a minimum for synthesis purposes.

Trnka, whose corpus is the pronouncing dictionary by Daniel Jones, found 3,178 different monosyllabic words. It must be assumed that the disyllabic and polysyllabic words of the dictionary contain syllables which do not occur as separates, i.e. as monosyllabic words, so that the total number of syllables in this corpus is considerably higher.

Dewey found, in 100,000 words of printed material, 4,400 different syllables. 1,370 of these occurred more than 10 times each, and accounted for 93% of the total of the corpus.

One could probably use Dewey's total figure, 4,400, as an indication of the number of syllable-size segments required for speech synthesis. Assuming that all of the 7 intonation contours can occur on any syllable, the total number of segments for synthesis will be

$$4{,}400 \times 7 = 30{,}800$$

or, if only the most common syllables are considered necessary,

$$1{,}370 \times 7 = 9{,}590$$

6. Syllable dyads

The concept of the dyad may be extended to the syllable, in order to subsume the dynamics involved in the transition from syllable to syllable. The cuts would then be made in the relatively sustained portion of the nuclei, assuming that one can find such a sustained part (cf. 4.3), and no other cuts would be required. All segments would be of the general type $\frac{N}{2}\,M\,\frac{N}{2}$. Zero nucleus, as well as zero margin, would have to be added to our list of phonetic segments. Table 4 shows the principles for synthesizing utterances of various types. N_0 and M_0 represent zero nucleus and zero margin.

The maximum number of segments required could be found by multiplying the number of final half-syllables with the number of initial half-syllables. We have found that 938 NM sequences may be required for word-final position, and 713 MN sequences for word-initial position (4.1, Table 3); for the sake of the present argument, we shall disregard the additional MN and NM sequences which may be needed for word-medial position. To these sequences must be added combinations between margins and zero nucleus. Assuming that any margin may combine with zero nucleus, we arrive at the following figures (cf 3.1, table 2):

NM (including NM_0)	938	MN (including M_0N)	713
N_0M	168	MN_0	66
NM total	1,106	MN total	779

$$NM \times MN = 1{,}106 \times 779 = 861{,}574$$

It seems hardly likely that all these combinations actually occur in English, but there are no data available which would indicate how many segments would be needed for synthesis.

It appears probable, though, that the number of syllable dyads would be greater than the number of syllables, since there are probably fewer general restrictions on permissible phoneme sequences in utterance chunks of more than syllable length than in syllable-long utterance chunks. Thus, linguists usually find that they can state the characteristic distribution of each phoneme more easily in relation to the syllable than to any larger phonological unit (Haugen, 1956). The complexity involved in stating the combinatory rules for phonemes in words of more than one syllable is apparent in Trnka's monograph.

Finally, one would have to multiply the number of $\frac{N}{2}\,M\,\frac{N}{2}$ sequences with the number of required prosodic conditions. Since the pitch levels and pitch glides seem to be associated with the syllable as their basic distributional unit, it would be necessary to multiply the number of NM sequences and the number of MN sequences separately with the required number of intonation contours.

$$NM = 1{,}106 \times 7 = 7{,}742$$
$$MN = 779 \times 7 = 5{,}453$$

The maximum number of syllable dyad segments required for speech synthesis would thus be $7{,}742 \times 5{,}453 = 42{,}217{,}126$

TABLE 4

UTTERANCE TYPE	SEGMENTS	EXAMPLE
N	$N_0 M_0 \frac{N}{2} + \frac{N}{2} M_0 N_0$	$\frac{\text{aɪ}}{2} + \frac{\text{aɪ}}{2}$ = [aɪ] I
MN	$N_0 M \frac{N}{2} + \frac{N}{2} M_0 N_0$	fr $\frac{\text{i}}{2} + \frac{\text{i}}{2}$ = [fri] free
NM	$N_0 M_0 \frac{N}{2} + \frac{N}{2} M N_0$	$\frac{\text{ɑ}}{2} + \frac{\text{ɑ}}{2}$ rm = [ɑrm] arm
MNM	$N_0 M \frac{N}{2} + \frac{N}{2} M N_0$	skr $\frac{\text{æ}}{2} + \frac{\text{æ}}{2}$ tʃ = [skrætʃ] scratch
NMN	$N_0 M_0 \frac{N}{2} + \frac{N}{2} M \frac{N}{2} + \frac{N}{2} M_0 N_0$	$\frac{\text{i}}{2} + \frac{\text{i}}{2}$ z $\frac{\text{i}}{2} + \frac{\text{i}}{2}$ = [izi] easy
MNNM	$N_0 M \frac{N}{2} + \frac{N}{2} M_0 \frac{N}{2} + \frac{N}{2} M N_0$	b $\frac{\text{aɪ}}{2} + \frac{\text{aɪ}}{2} \frac{\text{ɪ}}{2} + \frac{\text{ɪ}}{2}$ ŋ = [baɪɪŋ] buying
MNMNMNM	$N_0 M \frac{N}{2} + \frac{N}{2} M \frac{N}{2} + \frac{N}{2} M \frac{N}{2} + \frac{N}{2} M N_0$	l $\frac{\text{ɪ}}{2} + \frac{\text{ɪ}}{2}$ ŋgw $\frac{\text{ɪ}}{2} + \frac{\text{ɪ}}{2}$ st $\frac{\text{ɪ}}{2} + \frac{\text{ɪ}}{2}$ ks = [lɪŋgwɪstɪks] linguistics

Examples showing the principles for synthesizing utterances of various types from syllable dyads.

It is fairly obvious that this figure is by far too high as a minimum for synthesis of speech with relatively high intelligibility. However, there does not at present seem to be any way to decide just how many segments would be needed, except by undertaking some large-scale phonemic transcription of running texts, with frequency counts of the various syllable dyads.

7. WORDS

The longest phoneme sequence which could usefully serve as the basic segment in speech synthesis is probably the word. The term is here used in its traditional sense. Homonyms would for our purposes be the same word, whereas synonyms would be different words. Forms like /rid/ *read,* /ridz/ *reads,* /rɛd/ *read,* /ridɪŋ/ *reading* would also constitute different words. Thus, two linguistic forms of word-length are different words if they differ in phonemic composition.

If words are accepted as segments for speech synthesis, it will hardly be necessary to store as many segments as there are entries in, for example, the *Oxford English Dictionary.* One would have to base one's inventory on some kind of frequency count.

In 100,000 words of printed material Dewey found 10,119 different words. Out of these, 1,027 words occur more than 10 times each, and make up 78·6% of the total corpus.

In 80,000 words of telephone conversations, French *et al.* found 2,822 different words. When inflected forms were not counted as separate words, they found 2,240 different words. Out of these 2,240 words, 737 occur in at least 1% of the 1,900 telephone conversations, and together they constitute 96% of the total occurrence of words.

In 288,152 words of college classroom speeches, Black and Ausherman (1955) found 6,826 different words. However, the number of phonemically different forms of word-length in their data is greater, since they considered related inflected and derivational forms, and even homographs with different phonemic content, to be the same word.

Other word frequency studies are found, for example, in publications by Fossum (1944), Fraprie (1950), Thorndike (1931 and 1944), and Voelker (1942). Dale has published a bibliography of vocabulary studies (1949).

The number of segments required depends on the use to which the synthesizer will be put. One could make separate word frequency counts for each type of material one would expect to synthesize. Thus, if the synthesizer was to be used only in sending messages to, say, the operators of space ships, one would probably need a considerably smaller inventory than what would be required for the transmission of learned scientific material of varied nature, or for the exchange of ideas on cultural topics. It is not unlikely, however, that the number of words would be smaller than the number of syllable dyads.

In any event, the number of words would have to be multiplied by the number of prosodic conditions required. Assuming that disyllabic and polysyllabic words would be stressed in one way only, i.e. disregarding potential differences in degree of stress on the various syllables and in stress placement, one would have to consider more than seven intonation patterns for each word of more than one syllable. For the syllable which can potentially carry a gliding intonation contour, i.e. the potentially stressed syllable, has seven possible pitch patterns, but each of these can be combined with different pitch levels on each of the remaining syllables. Also, some words might have more than one potentially stressed syllable, and thus more than one syllable which can take any of the seven essential intonation contours. Thus, the number of words might have to be multiplied by a relatively high number of prosodic conditions.

8. Summary and conclusion

The preceding pages give a survey of the types of segment that might be useful in a form of speech synthesis where the building blocks are successive segments in time. There is also an estimate of the number of segments that might have to be stored for each type of synthesis.

Table 5 is a summary of these estimates. " Phoneme sequences ", in the second column, gives a suggested minimum for the number of segmental phoneme sequences that might have to be represented in the inventory. When the segment is of phoneme length, the phoneme sequence consists of one phoneme only. The column " Segments " suggests a minimum for the total segment inventory when various prosodic conditions are taken into account.

The figures are tentative. They suggest only the approximate and relative size of the inventory for each type of segment. First, the data are incomplete on many points. Thus, we know little about which particular phoneme sequences may occur

TABLE 5

SEGMENT TYPE	PHONEME SEQUENCES	SEGMENTS
Phoneme	37	155
Phoneme dyad	1,218	8,460
IC of syllable	226	650
Half-syllable	1,651	11,557
Syllable:		
Dewey, total	4,400	30,800
Dewey, most frequent	1,370	9,590
Syllable dyad	861,574 (?)	42,217,126 (?)
Word:		
Dewey, total	10,119	
Dewey, most frequent	1,027	
French, total	2,822	(?)
French, most frequent	737	
Black and Ausherman	6,826	

Number of segments required for speech synthesis from various segment types.

utterance-medial position. Second, the size of the inventory depends on considerations of the frequency of occurrence of the various units in the particular type of material that the machine will be required to synthesize. Third, the estimates are in part based on certain hypotheses about the nature of speech ; it would be necessary to test in actual synthesis, for each type of segment, the necessity for increasing the inventory or the possibility of decreasing it.

At the present one can give only a theoretical evaluation and comparison of the various types of synthesis.

(a) When we compare the figures of Table 5, one thing stands out: With the exception of the phoneme dyad and the syllable dyad, the size of the inventory increases with the length of the segment. Thus, there is a *direct* relationship between the length of the segment and the size of the inventory. This could be predicted for a system like language, which is characterized by the fact that a relatively small number of units on one hierarchical level combine in various ways to form a large number of units on a higher level of the hierarchy.

The relationship obtains only when the segments correspond to some linguistic unit. The segment inventory becomes disproportionately large when the segments are not co-terminous with linguistic units. This is the case for phoneme dyads and for syllable dyads.

(b) From the point of view of intelligibility, however, it is better to use as long segments as possible. When we segment into an utterance in order to establish invariable building blocks for speech synthesis, we interfere with the natural dynamics of speech, with the continuously changing spectrum and fundamental frequency. If some of these dynamics of speech could be incorporated in the segments, it seems likely that intelligibility would be higher.

(c) With present techniques, the number of units to be stored is probably a minor problem. Thus, if our primary objective is to establish an efficient communication system, we would no doubt select one of the longer types of segment as basis.

However, speech synthesis is today an important research tool for speech analysis. (Peterson and Sivertsen, 1960; Stevens, 1960). Testing methodically the various types of segment considered above, making systematic changes in the inventory for each type, disregarding or including specified phonetic or phonemic features (duration, stress, juncture, pitch phenomena, allophonic variations of phonemes, changes in the spectrum during a phoneme-long segment, etc.), might throw new light on many aspects of speech and language. By pitting synthesis from short segments against synthesis from longer segments, one could study the relative importance, for perception, of the dynamics of speech, on the one hand, and of the target values or invariances, on the other.

Speech synthesis from stored segments provides a means of checking certain theories about linguistic structure, and phonetic statements about the nature of speech. For English one could, for example, test the hypothesis that intonation patterns are combinations of a limited number of pitch levels, or that 2 (or 3, or 4) contrastive stresses are needed in order to account for all stress differences, or that internal open juncture contrasts with absence of juncture. One could further test statements about the phonetic nature of stress contrasts or juncture contrasts, and check descriptions of the invariance of phonemes, on the one hand, and of their allophonic variations, on the other. More generally, such synthesis may be used in studies of the basic problem of segmentation. It is well known that there are a number of difficulties involved in segmentation of the speech continuum. Doubts have been raised whether such segmentation is possible at all, acoustically or physiologically, except in a purely physical sense. X-ray studies of the physiological mechanism and acoustical analyses of the speech wave show continuously changing formations and patterns in many cases where an intuitive auditory-phonetic analysis finds a segment border. Speech synthesis might provide an answer to some of the problems.

The value of speech synthesis from stored segments as a research tool in speech analysis, like that of other types of speech synthesis, lies less in the new and original data on language and speech that it might provide: it is of importance first and foremost as a check on data and hypotheses derived by other means, such as traditional linguistic analysis and instrumental studies of the vocal mechanism and of the speech wave.

References

Black, J. W., and Ausherman, M. (1955). The Vocabulary of College Students in Classroom Speeches (Bureau of Educational Research, Ohio State University).

Bloomfield, L. (1933). Language (New York).

Bolinger, D. L. (1958). A theory of pitch accent in English. *Word*, 14, 109.

Dale, E. (1949). Bibliography of Vocabulary Studies (Bureau of Educational Research, Ohio State University).

DEWEY, G. (1923). Relativ Frequency of English Speech Sounds (Cambridge, Mass.).

FAIRBANKS, G., EVERITT, W. L., and JAEGER, R. P. (1954). Method for time or frequency compression-expansion of speech. *Trans. IRE-PGA*, AU2, 7.

FOSSUM, E. C. (1944). An analysis of the dynamic vocabulary of junior college students. *Speech Monographs*, 40, 88.

FRAPRIE, F. R. (1950). The twenty commonest English words . . . from a count of 242,423 words of English text taken from fifteen English authors and many newspapers. *World Almanac and Book of Facts for 1950*, 543 (World-Telegram, New York).

FRENCH, N. R., CARTER, C. W., and KOENIG, W., JR. (1930). The words and sounds of telephone conversations. *Bell Telephone System Technical Publications*, Monograph B-491.

FRY, D. B. (1958). Experiments in the perception of stress. *Language and Speech*, 1, 126.

FUNK AND WAGNALLS COMPANY (1925). New Standard Dictionary of the English Language (New York and London).

GLEASON, H. A., JR. (1955). An Introduction to Descriptive Linguistics (New York).

HARRIS, C. M. (1953). A study of the building blocks of speech. *J. acoust. Soc. Amer.*, 25, 962.

HAUGEN, E. (1956). The syllable in linguistic description. *For Roman Jakobson*, 213.

HILL, A. A. (1958). Introduction to Linguistic Structures (New York).

HOCKETT, C. F. (1955). A Manual of Phonology (Indiana University Publications in Anthropology and Linguistics, also Memoir 11 of the *International Journal of American Linguistics*, Baltimore).

HOCKETT, C. F. (1958). A Course in Modern Linguistics (New York).

JONES, D. (1956). An English Pronouncing Dictionary. 11th ed. (London and New York).

KENYON, J. S., and KNOTT, T. A. (1944). A Pronouncing Dictionary of American English (Springfield, Mass.).

LEHISTE, I. (1959). An Acoustic-Phonetic Study of Internal Open Juncture (Speech Research Laboratory, University of Michigan, Report No. 2, Ann Arbor).

LEHISTE, I., and PETERSON, G. E. (1960). Diphthongs, glides, and transitions. To appear in *J. acoust. Soc. Amer.*

LIBERMAN, A. M., *et al.* (1959). Minimal rules for synthesizing speech. *J. acoust. Soc. Amer.*, 31, 1490.

MALONE, K. (1936). The phonemic structure of English monosyllables. *American Speech*, 11, 205.

MALONE, K. (1959). The phonemes of current English. *Studies in Heroic Legend and Current Speech*, ed. S. Einarsson and N. E. Eliason (Copenhagen).

O'CONNOR, J. D., and TRIM, J. L. M. (1953). Vowel, consonant, and syllable—a phonological definition. *Word*, 9, 103.

PETERSON, G. E. (1958). Some observations on speech. *Quart. J. Speech*, 44, 402.

PETERSON, G. E., and BARNEY, H. L. (1952). Control methods used in a study of the vowels. *J. acoust. Soc. Amer.*, 24, 175.

PETERSON, G. E., and SIVERTSEN, E. (1960). Objectives and techniques of speech synthesis. *Language and Speech*, 3, 84.

PETERSON, G. E., WANG, W. S-Y., and SIVERTSEN, E. (1958). Segmentation techniques in speech synthesis. *J. acoust. Soc. Amer.*, 30, 739.

PIKE, K. L. (1945). The Intonation of American English (Ann Arbor, Michigan).

SIVERTSEN, E. (1960). Cockney Phonology (Oslo).

STEVENS, K. N. (1960). Toward a model for speech recognition. *J. acoust Soc. Amer.*, 32, 47.

THORNDIKE, E. L. (1931). A Teacher's Word Book of Twenty Thousand Words Found Most Frequently and Widely in General Reading for Children and Young People (Bureau of Publications, Teacher's College, Columbia University, New York).

THORNDIKE, E. L., and LORGE, I. (1944). The Teacher's Word Book of 30,000 Words. (Bureau of Publications, Teacher's College, Columbia University, New York).

TRAGER, G. L., and SMITH, H. L., JR. (1951). An Outline of English Structure (Norman, Oklahoma).

TRNKA, B. (1935). A Phonological Analysis of Present-Day Standard English (Prague).

TRUBY, H. M. (1959). Acoustic-Cineradiographic Analysis Considerations with Especial Reference to Certain Consonantal Complexes, Supplement to *Acta Radiologica*, (Stockholm).

VOELKER, C. H. (1942). The one-thousand most frequent spoken-words. *Quart. J. Speech*, **28**, 189.

WALLACE, B. J. (1950). A Quantitative Analysis of Consonant Clusters in Present-Day English (Unpublished University of Michigan dissertation, Ann Arbor).

WANG, W. S-Y., and CRAWFORD, J. (1960). Frequency studies of English consonants. *Language and Speech*, **3**, 131.

WANG, W. S-Y., and PETERSON, G. E. (1958). Segment inventory for speech synthesis. *J. acoust. Soc. Amer.*, **30**, 743.

WHORF, B. L. (1940). Linguistics as an exact science. *The Technology Review*, **42**, No. 6.

WOOD, C. (1943). Wood's Unabridged Rhyming Dictionary (Cleveland and New York).

APPENDIX

The following is a list of (1) all the word-initial and word-final margins (M) which have been documented for the GA and RP dialects of English, (2) the particular syllable nuclei (N) with which each margin may combine, and (3) key-words for these MN and NM combinations. The list represents the " maximum " of 2.31.

The column MIN. (minimum) indicates whether the particular phoneme combination is adopted in the " suggested minimum " of 2.32.

The data were subjected to certain phonemic normalizations: 2.2.

Variant pronunciations are indicated, and a word may be listed as key-word for different phoneme sequences. Thus, *pierce* is the key-word for /-irs/ as well as for /-ɪrs/, and *lunch* is given as example of both /-əntʃ/ and /-ənʃ/. Two key-words are sometimes given, in order to account for the distribution of RP /æ, ɑ, ɒ, ɔ/ as well as of GA /æ, ɑ, ɔ/. NM sequences which occur only in dialects where /r/ is not found before consonants and junctures (" r-less dialects ") are enclosed between parentheses.

Some sequences are rare and seem to be characteristic of certain dialects or idiolects only. In such cases the source for the key-word is indicated in parentheses. The following abbreviations are used:

FW	New Standard Dictionary of the English Language, Funk and Wagnalls Company.
Hi	A. A. Hill, Introduction to Linguistic Structures.
Ho	C. F. Hockett, A Course in Modern Linguistics.
J	D. Jones, An English Pronouncing Dictionary.
K	J. S. Kenyon and T. A. Knott, A Pronouncing Dictionary of American English.
M	K. Malone, The phonemic structure of English monosyllables.

M 1959	The phonemes of current English.
T	B. Trnka, A Phonological Analysis of Present-Day Standard English.
Wh	B. L. Whorf, Linguistics as an exact science.
Wo	C. Wood, Wood's Unabridged Rhyming Dictionary.
r-less	occurs only in so-called r-less dialects, i.e. those dialects where /r/ never occurs before consonants or junctures.

Notes are added for those phoneme combinations that are excluded from the suggested minimum, giving reasons for their omission:

Rare : occurs only in rare words, for example, scientific or otherwise specialized terms (example: /pt/ in *ptomaine*), or represents an unusual pronunciation (example: /ps/ in *psycho-*).
Foreign : occurs only in foreign names or words which have not been assimilated into English (examples: /pʃ/ in *Przemysl*, /pf/ in *pfennig*). The decision as to whether a certain MN or NM combination is foreign to English or not, is necessarily arbitrary. From another point of view, such combinations might be considered " rare ".
Not GA : occurs only in dialects other than the particular type of GA under consideration in this study (example: /tj/ in *tune*).
Only ɚpts ? (ɚθt, etc.): /pts/ (/θt/, etc.) probably occurs after /r/ only if [ɚC] is interpreted as /ərC/.
>: can be replaced by, without appreciable loss of intelligibility. This applies to some unusual sequences which are generally paralleled by some other sequences in other idiolects (example: /ps/ > /s/ in *psycho-*), sequences which are not used in GA (example: /tj/ > /t/ in *tune*), sequences which are freely varying with other sequences, the distribution being partly idiolectally conditioned (example: /nʃ/ > /ntʃ/ in *lunch*), and sequences which are often replaced by other sequences in normal colloquial style (example: /tθs/ > /ts/ in *eighths*).

I. Initial MN Combinations

ONSET	NUC.	KEY-WORD	MIN.	NOTES
zero	i	eat	+	
	ɪ	it	+	
	ɛ	edge	+	
	æ	at	+	
	ɑ	ox, art	+	
	ɔ	ought	+	
	ʊ	umlaut, oomph	+	
	u	ooze	+	
	ə	up	+	
	ɚ	earth	+	
	eɪ	eight	+	
	aɪ	ice	+	
	aʊ	out	+	
	ɔɪ	oil	+	
	oʊ	own	+	
p	i	pea	+	
	ɪ	pit	+	
	ɛ	pet	+	
	æ	pat	+	
	ɑ	pot, part	+	
	ɔ	pause	+	
	ʊ	put	+	
	u	pool	+	
	ə	puff	+	
	ɚ	pearl	+	
	eɪ	pay	+	
	aɪ	pie	+	
	aʊ	pound	+	
	ɔɪ	poise	+	
	oʊ	poke	+	
pt	ɛ	pterosaur (J)	–	Rare, > t
pf	ɛ	pfennig (J, K)	–	Foreign
ps	ɪ	psittachosis (T)	–	Rare, > s
	ɑ	psoriasis (T)	–	
	aɪ	psycho- (T)	–	
psj	u	pseudo- (T)	–	Rare, > s
pʃ	ɛ	Przemysl (J)	–	Foreign, Rare
	ɔ	pshaw (J)	–	
	u	two	+	
	ə	ton	+	
	ɚ	turn	+	
	eɪ	tame	+	
	aɪ	time	+	
	aʊ	towel	+	
	ɔɪ	toy	+	
	oʊ	tone	+	
ts	ɪ	Tsimshian (Ho)	–	Rare, > t, z
	ɛ	tsetse (T)	–	
	ɑ	czar (T)	–	
tʃ	i	cheese	+	
	ɪ	chip	+	
	ɛ	check	+	
	æ	chap	+	
	ɑ	chop, chart	+	
	ɔ	chalk	+	
	ʊ	Chunking	+	
	u	choose	+	
	ə	chunk	+	
	ɚ	church	+	
	eɪ	chase	+	
	aɪ	chide	+	
	aʊ	chow	+	
	ɔɪ	choice	+	
	oʊ	chose	+	
tm	i	tmesis	–	Rare
tɬ	ɪ	Tlingit (Ho)	–	Rare
tr	i	tree	+	
	ɪ	trip	+	
	ɛ	treck	+	
	æ	track	+	
	ɑ	trot, trance	+	
	ɔ	trawl	+	
	ʊ	Truro	+	
	u	true	+	
	ə	truck	+	
	eɪ	tray	+	

ONSET	NUC.	KEY-WORD	MIN.	NOTES
pl	i	please	+	
	ɪ	plinth	+	
	ɛ	pledge	+	
	æ	plan	+	
	ɑ	plot, plant	+	
	ɔ	plausible	+	
	ʊ	plural	+	
	u	plume	+	
	ə	plump	+	
	eɪ	plane	+	
	aɪ	ply	+	
	aʊ	plough	+	
	ɔɪ	ploy	+	
	oʊ	plosion	+	
pr	i	preach	+	
	ɪ	print	+	
	ɛ	press	+	
	æ	prank	+	
	ɑ	prop, prance	+	
	ɔ	prawn	+	
	u	prune	+	
	ə	Prussia	+	
	eɪ	prate	+	
	aɪ	price	+	
	aʊ	proud	+	
	oʊ	probe	+	
pj	æ	piano	−	pjæ > pɪæ
	ɔ	pure (J)	−	pjɔr > pjʊr
	ʊ	pure	+	
	u	pew	+	
pw	ɛ	pueblo	−	Foreign, > pu, pʊ
t	i	tea	+	
	ɪ	tip	+	
	ɛ	test	+	
	æ	tap	+	
	ɑ	top, tar	+	
	ɔ	talk	+	
	ʊ	took	+	

ONSET	NUC.	KEY-WORD	MIN.	NOTES
	aɪ	try	+	
	aʊ	trout	+	
	ɔɪ	Troy	+	
	oʊ	troll	+	
tj	ʊ	Turin	−	Not GA
	u	tune	—	
tw	i	tweezer	+	
	ɪ	twin	+	
	ɛ	twenty	+	
	æ	twang	+	
	ɑ	'twas	+	
	ʊ	'twould	+	
	ɚ	twirl	+	
	eɪ	twain	+	
	aɪ	twice	+	
k	i	key	+	
	ɪ	kin	+	
	ɛ	kept	+	
	æ	cap	+	
	ɑ	cop, cart	+	
	ɔ	caught	+	
	ʊ	cook	+	
	u	cool	+	
	ə	cup	+	
	ɚ	curl	+	
	eɪ	cake	+	
	aɪ	kite	+	
	aʊ	cow	+	
	ɔɪ	coy	+	
	oʊ	coke	+	
kn	aʊ	knout (T)	−	Rare, > n
kl	i	clean	+	
	ɪ	click	+	
	ɛ	cleanse	+	
	æ	clan	+	
	ɑ	clot, Clark	+	
	ɔ	clause	+	
	u	clue	+	

ONSET	NUC.	KEY-WORD	MIN.	NOTES
	ə	club	+	
	ɚ	clerk	+	
	eɪ	clay	+	
	aɪ	climb	+	
	aʊ	cloud	+	
	ɔɪ	cloy	+	
	oʊ	close	+	
kr	i	cream	+	
	ɪ	crisp	+	
	ɛ	cress	+	
	æ	cram	+	
	ɑ	crop, craft	+	
	ɔ	crawl	+	
	ʊ	crook	+	
	u	crude	+	
	ə	crumb	+	
	eɪ	crave	+	
	aɪ	cry	+	
	aʊ	crowd	+	
	ɔɪ	Croydon	+	
	oʊ	crow	+	
kj	ɔ	cure (J)	−	kjɔr > kjʊr
	ʊ	cure	+	
	u	queue	+	
kw	i	queen	+	
	ɪ	quick	+	
	ɛ	quest	+	
	æ	quack	+	
	ɑ	quad, quaff	+	
	ɔ	qualm	+	
	ə	quadrille	+	
	ɚ	quirk	+	
	eɪ	quake	+	
	aɪ	quite	+	
	ɔɪ	quoit	+	
	oʊ	quote	+	
b	i	beat	+	
	ɪ	bit	+	
	aɪ	bright	+	
	aʊ	brown	+	
	ɔɪ	broil	+	
	oʊ	broke	+	
bj	ɔ	bureau (J)	−	bjɔr > bjʊr
	ʊ	bureau	+	
	u	bugle	+	
bw	ɛ	bueno	−	Foreign, > bu, bʊ
	ɑ	bwana	−	
d	i	deep	+	
	ɪ	dip	+	
	ɛ	depth	+	
	æ	damp	+	
	ɑ	dock, dart	+	
	ɔ	daub	+	
	ʊ	dour	+	
	u	doom	+	
	ə	duck	+	
	ɚ	dirt	+	
	eɪ	day	+	
	aɪ	die	+	
	aʊ	down	+	
	ɔɪ	doily	+	
	oʊ	dough	+	
dv	i	Dvina	−	Foreign
	ɔ	Dvorak	−	
dʒ	i	Jean	+	
	ɪ	Jim	+	
	ɛ	gem	+	
	æ	jam	+	
	ɑ	jot, jar	+	
	ɔ	George	+	
	ʊ	jury	+	
	u	Jew	+	
	ə	jump	+	
	ɚ	jerk	+	
	eɪ	Jane	+	
	aɪ	gibe	+	

ONSET	NUC.	KEY-WORD	MIN.	NOTES
	ɛ	bet	+	
	æ	bat	+	
	ɑ	box, barn	+	
	ɔ	bought	+	
	ʊ	book	+	
	u	boot	+	
	ə	buck	+	
	ɚ	birch	+	
	eɪ	bait	+	
	aɪ	bite	+	
	aʊ	bound	+	
	ɔɪ	boy	+	
	oʊ	boat	+	
bd	ɛ	bdellium (T)	−	Rare, > d
bl	i	bleak	+	
	ɪ	bliss	+	
	ɛ	bless	+	
	æ	black	+	
	ɑ	blot, blarney	+	
	ɔ	Blore (J)	+	
	u	bloom	+	
	ə	blood	+	
	ɚ	blur	+	
	eɪ	blame	+	
	aɪ	blight	+	
	aʊ	blouse	+	
	oʊ	blow	+	
br	i	breed	+	
	ɪ	brim	+	
	ɛ	bread	+	
	æ	brand	+	
	ɑ	bronze, branch	+	
	ɔ	brought	+	
	ʊ	brook	+	
	u	brute	+	
	ə	brush	+	
	ɚ	Brer (J)	−	Rare
	eɪ	brake	+	

ONSET	NUC.	KEY-WORD	MIN.	NOTES
	aʊ	jowl	+	
	ɔɪ	joy	+	
	oʊ	joke	+	
dn	i	Dnieper (J)	−	Foreign
dr	i	dream	+	
	ɪ	drip	+	
	ɛ	dress	+	
	æ	drab	+	
	ɑ	drop, drama	+	
	ɔ	draw	+	
	ʊ	Drury Lane	+	
	u	drew	+	
	ə	drum	+	
	eɪ	drake	+	
	aɪ	dry	+	
	aʊ	drown	+	
	ɔɪ	Droitwich (J)	+	
	oʊ	drove	+	
dj	ɔ	during (J)	−	Not GA
	ʊ	during	—	
	u	dew	—	
dw	i	Dwina (J)	−	Rare
	ɪ	dwindle	+	
	ɛ	dwell	+	
	æ	dwang (M)	+	
	ɔ	dwarf	+	
	eɪ	dwale	+	
	aɪ	Dwight	+	
g	i	geese	+	
	ɪ	give	+	
	ɛ	get	+	
	æ	gap	+	
	ɑ	got, garden	+	
	ɔ	gaudy	+	
	ʊ	good	+	
	u	goose	+	
	ə	gun	+	
	ɚ	girl	+	

ONSET	NUC.	KEY-WORD	MIN.	NOTES
	eɪ	gave	+	
	aɪ	guy	+	
	aʊ	gown	+	
	ɔɪ	goiter	+	
	oʊ	go	+	
gn	aɪ	gneiss (Hi)	–	Rare, > n
	oʊ	gnosis (T)	–	
gl	i	glee	+	
	ɪ	glib	+	
	ɛ	glen	+	
	æ	glad	+	
	ɑ	glottal, glass	+	
	ɔ	gloss	+	
	ʊ	Gluck	–	Foreign
	u	gloom	+	
	ə	glove	+	
	eɪ	glade	+	
	aɪ	glide	+	
	aʊ	glout (M)	+	
	oʊ	glow	+	
gr	i	green	+	
	ɪ	grin	+	
	ɛ	Gregory	+	
	æ	grand	+	
	ɑ	grotto, grass	+	
	ɔ	grog	+	
	ʊ	gruel (J)	+	
	u	grew	+	
	ə	grunt	+	
	eɪ	grey	+	
	aɪ	gripe	+	
	aʊ	ground	+	
	ɔɪ	groin	+	
	oʊ	grow	+	
gj	u	gules	+	
gw	ɪ	Gwynn	+	
	ɛ	Gwendolyn	+	
	ɑ	guano	–	Foreign, > gu, gʊ

ONSET	NUC.	KEY-WORD	MIN.	NOTES
	ʊ	fructose	+	
	u	fruit	+	
	ə	front	+	
	eɪ	freight	+	
	aɪ	fright	+	
	aʊ	frown	+	
	ɔɪ	Freud	+	
	oʊ	froze	+	
fj	ɔ	fiord	–	> fɪ
	u	few	+	
θ	i	theme	+	
	ɪ	thin	+	
	ɛ	theft	+	
	æ	thank	+	
	ɑ	tharm (M)	+	
	ɔ	thaw	+	
	ʊ	Thuringia (K)	+	
	u	Thule	+	
	ə	thumb	+	
	ɚ	thirst	+	
	eɪ	thane	+	
	aɪ	thigh	+	
	aʊ	thousand	+	
	oʊ	thole	+	
θr	i	three	+	
	ɪ	thrift	+	
	ɛ	threat	+	
	æ	thrash	+	
	ɑ	throb	+	
	ɔ	thrall	+	
	u	through	+	
	ə	thrust	+	
	eɪ	thrave	+	
	aɪ	thrive	+	
	aʊ	throve	+	
θj	ʊ	thurible (K)	–	Not GA
	u	thew	–	
θw	ɪ	Thwing (M)	–	Rare

ONSET	NUC.	KEY-WORD	MIN.	NOTES	ONSET	NUC.	KEY-WORD	MIN.	NOTES
	aɪ	guaiac	—			æ	thwack	+	
f	i	fiend	+			ɔ	thwart	+	
	ɪ	fin	+			eɪ	thwaite	+	
	ɛ	fence	+		s	i	see	+	
	æ	fan	+			ɪ	sit	+	
	ɑ	fop, far	+			ɛ	set	+	
	ɔ	fought	+			æ	sat	+	
	ʊ	foot	+			ɑ	sock, sardine	+	
	u	food	+			ɔ	sought	+	
	ə	fun	+			ʊ	soot	+	
	ɚ	fur	+			u	suit	+	
	eɪ	fame	+			ə	son	+	
	aɪ	find	+			ɚ	sir	+	
	aʊ	found	+			eɪ	say	+	
	ɔɪ	foist	+			aɪ	sigh	+	
	oʊ	foe	+			aʊ	sound	+	
fθ	aɪ	phthisis	—	Rare		ɔɪ	soya	+	
fl	i	flea	+			oʊ	so	+	
	ɪ	flip	+		sp	i	speak	+	
	ɛ	flesh	+			ɪ	spit	+	
	æ	flash	+			ɛ	speck	+	
	ɑ	flop, Flanders	+			æ	spat	+	
	ɔ	flaw	+			ɑ	spot, spark	+	
	ʊ	fluoroscope	+			ɔ	spawn	+	
	u	flue	+			ʊ	spoor	+	
	ə	flood	+			u	spoon	+	
	ɚ	flirt	+			ə	spud	+	
	eɪ	flame	+			ɚ	spur	+	
	aɪ	fly	+			eɪ	spade	+	
	aʊ	flout	+			aɪ	spy	+	
	ɔɪ	Floyd	+			aʊ	spout	+	
	oʊ	flow	+			ɔɪ	spoil	+	
fr	i	free	+			oʊ	spoke	+	
	ɪ	frill	+		spl	i	spleen	+	
	ɛ	fresh	+			ɪ	split	+	
	æ	frank	+			ɛ	splendid	+	
	ɑ	from, France	+			æ	splash	+	
	ɔ	fraud	+			ɑ	splotch	+	

ONSET	NUC.	KEY-WORD	MIN.	NOTES
	ə	splutter	+	
	ɚ	splurge	+	
	eɪ	splay	+	
	aɪ	splice	+	
spr	i	spree	+	
	ɪ	spring	+	
	ɛ	spread	+	
	æ	sprat	+	
	ɑ	sprocket	+	
	ɔ	sprawl	+	
	u	spruce	+	
	ə	sprung	+	
	eɪ	spray	+	
	aɪ	spry	+	
	aʊ	sprout	+	
	oʊ	Sproule (J)	−	Rare
spj	ʊ	spurious	+	
	u	spew	+	
st	i	steam	+	
	ɪ	stick	+	
	ɛ	step	+	
	æ	stack	+	
	ɑ	stop, start	+	
	ɔ	stalk	+	
	ʊ	stood	+	
	u	stoop	+	
	ə	stuck	+	
	ɚ	stir	+	
	eɪ	stay	+	
	aɪ	sty	+	
	aʊ	stout	+	
	oʊ	stoke	+	
str	i	street	+	
	ɪ	string	+	
	ɛ	stress	+	
	æ	strap	+	
	ɑ	strop, strata (J)	+	
	ɔ	straw	+	

ONSET	NUC.	KEY-WORD	MIN.	NOTES
	ɪ	squint	+	
	ɛ	squelch	+	
	ɑ	squat	+	
	ɔ	squaw	+	
	ɚ	squirt	+	
	eɪ	squama	−	Rare
	aɪ	squire	+	
sg	ɑ	sgalag (FW)	−	Rare, > sk
sf	i	sphere	−	ir > ɪr
	ɪ	sphinx	+	
	ɛ	spherical	+	
	æ	sphagnum (J)	−	Rare
	ɔ	sforzando	−	Rare
	u	sfumato (FW)	−	Rare
	oʊ	sfogato (FW)	−	Rare
sfr	ɪ	sphrigosis (FW)	−	Rare
	æ	sphragide (FW)	−	
	ə	sphragistic (K)	−	
	eɪ	sfregazzi (FW)	−	
sθ	i	sthenia (FW)	−	Rare
	ɪ	sthenia (FW)	−	
	ɛ	sthenic	−	
sv	ɛ	svelte	−	> sf
	ɑ	svarabhakti	−	
sm	i	smeek (FW)	+	
	ɪ	smith	+	
	ɛ	smell	+	
	æ	smash	+	
	ɑ	smock, smart	+	
	ɔ	small	+	
	u	smooth	+	
	ə	smug	+	
	ɚ	smirk	+	
	eɪ	Smale (J)	+	
	aɪ	smile	+	
	oʊ	smoke	+	
smj	u	smew	−	Rare
sn	i	sneak	+	

ONSET	NUC.	KEY-WORD	MIN.	NOTES
	ʊ	strook	+	
	u	strew	+	
	ə	struck	+	
	eɪ	stray	+	
	aɪ	strike	+	
	aʊ	Stroud	+	
	ɔɪ	stroy (K)	+	
	oʊ	stroke	+	
stj	u	stupid	−	Not GA
sk	i	ski	+	
	ɪ	skid	+	
	ɛ	sketch	+	
	æ	scan	+	
	ɑ	Scot, scar	+	
	ɔ	scaup	+	
	u	scoop	+	
	ə	skull	+	
	ɚ	skirt	+	
	eɪ	skate	+	
	aɪ	sky	+	
	aʊ	scout	+	
	oʊ	scope	+	
skl	ɪ	sclere	−	Rare
	ɛ	sklent (M 1959)	−	
	æ	sclaff	−	
skr	i	scream	+	
	ɪ	script	+	
	æ	scratch	+	
	ɑ	scrod	+	
	ɔ	scrawl	+	
	u	screw	+	
	ə	scrub	+	
	eɪ	scrape	+	
	aɪ	scribe	+	
	aʊ	scrounge	+	
	oʊ	scroll	+	
skj	u	skew	+	
skw	i	squeeze	+	

ONSET	NUC.	KEY-WORD	MIN.	NOTES
	ɪ	sniff	+	
	ɛ	snell	+	
	æ	snatch	+	
	ɑ	snob, snarl	+	
	ɔ	snort	+	
	ʊ	snooks (Wo)	−	Rare
	u	snoop	+	
	ə	snub	+	
	ɚ	Snerd	−	Rare
	eɪ	snail	+	
	aɪ	snipe	+	
	aʊ	snout	+	
	oʊ	snow	+	
snj	u	Snewin (M 1959)	−	Not GA
sl	i	sleet	+	
	ɪ	slip	+	
	ɛ	sled	+	
	æ	slack	+	
	ɑ	slot, slant	+	
	ɔ	slaw	+	
	u	sloop	+	
	ə	slum	+	
	ɚ	slur	+	
	eɪ	sleigh	+	
	aɪ	slight	+	
	aʊ	slough	+	
	ɔɪ	sloid	+	
	oʊ	slow	+	
sj	u	suit	−	Not GA
sw	i	sweet	+	
	ɪ	swing	+	
	ɛ	swell	+	
	æ	swank	+	
	ɑ	swop	+	
	ɔ	swore	+	
	u	swoon	+	
	ə	swung	+	
	ɚ	swirl	+	

ONSET	NUC.	KEY-WORD	MIN.	NOTES	ONSET	NUC.	KEY-WORD	MIN.	NOTES
	eɪ	sway	+			ə	thus	+	
	aɪ	swine	+			eɪ	they	+	
	aʊ	swound (K)	–	Rare		aɪ	thy	+	
	ɔɪ	Swoyersville (K)	–	Rare		aʊ	thou	+	
	oʊ	swollen	+			oʊ	though	+	
ʃ	i	sheet	+		z	i	z	+	
	ɪ	ship	+			ɪ	zip	+	
	ɛ	shelf	+			ɛ	zest	+	
	æ	shack	+			æ	zam (M)	+	
	ɑ	shock, sharp	+			ɑ	czar	+	
	ɔ	shawl	+			ʊ	zouave (K)	–	Foreign
	ʊ	shook	+			u	zoo	+	
	u	shoot	+			ə	Zutphen (K)	–	Foreign
	ə	shun	+			ɚ	zircon (K)	–	Rare
	ɚ	shirt	+			eɪ	zeta	+	
	eɪ	shake	+			aɪ	Zion	+	
	aɪ	shy	+			aʊ	zounds (J, K)	–	Rare
	aʊ	shout	+			oʊ	zone	+	
	oʊ	show	+		zbl	ə	'sblood (Hi)	–	Rare
ʃp		(names, Hi)	–	Foreign	zd	ɛ	'sdeath (T)	–	Rare
ʃt		(names, Hi)	–	Foreign	zg	ɑ	Sganarelle (FW)	–	Foreign
ʃk		(names, Hi)	–	Foreign	zl	ɑ	zloty (J, K)	–	Foreign
ʃm	ɪ	Schmidt	–	Foreign	zj	ʊ	Zürich	–	Foreign
	u	shmoo	–	Rare		u	Zeus	–	
ʃn	ɪ	schnitzel (FW)	–	Rare, Foreign	zw	ɪ	Zwingli	–	Foreign
	ɛ	schneller (FW)	–			ɑ	zouave (K)	–	
	æ	schnapps	–		ʒ	i	gigue (K)	–	Foreign
	ɑ	schnabel (FW)	–			ɪ	Giselle (Ho)	–	
	ʊ	shnook	–			æ	jalousie (J)	–	
	aɪ	schneider (FW)	–			ɑ	genre (K)	–	
	aʊ	schnouzer (K)	–			ɔ	jongleur (J)	–	
	oʊ	schnorrer (FW)	–			u	jupe (J)	–	
ʃl	i	Schliemann (K)	–	Rare, Foreign	m	i	meet	+	
	ɪ	Schlitz	–			ɪ	miss	+	
	ɛ	Schleswig (K)	–			ɛ	met	+	
	ɑ	schloss (FW)	–			æ	mat	+	
	ə	shlub	–			ɑ	mock, mart	+	
	eɪ	Schlegel (K)	–			ɔ	maw	+	

ONSET	NUC.	KEY-WORD	MIN.	NOTES
	ɪ	Schley (K)	—	
	aɪ	shriek	+	
ʃr	i	shrill	+	
	ɛ	shred	+	
	æ	shrank	+	
	ɑ	shroff (M)	+	
	u	shrewd	+	
	ə	shrug	+	
	aɪ	shrive	+	
	aʊ	shroud	+	
	oʊ	shrove	+	
ʃw	ɪ	Schwinn	—	Rare, Foreign
	ɛ	Schweppe (J)	—	
	ɑ	schwa	—	
	ɔ	Schwann	—	
v	i	veal	+	
	ɪ	victory	+	
	ɛ	vent	+	
	æ	van	+	
	ɑ	vox	+	
	ɔ	vault	+	
	u	voodoo	+	
	ə	vulgar	+	
	ɝ	verb	+	
	eɪ	vain	+	
	aɪ	vice	+	
	aʊ	vow	+	
	ɔɪ	voice	+	
	oʊ	vote	+	
vl	æ	Vladivostock	—	Foreign
vr	i	Vries (FW)	—	Foreign
	eɪ	vraisemblance (J)	—	
	aɪ	Vryburg (J)	—	
vj	u	view	+	
ð	i	thee	+	
	ɪ	this	+	
	ɛ	them	+	
	æ	that	+	

ONSET	NUC.	KEY-WORD	MIN.	NOTES
	ʊ	moor	+	
	u	mood	+	
	ə	mud	+	
	ɝ	mirk	+	
	eɪ	make	+	
	aɪ	my	+	
	aʊ	mound	+	
	ɔɪ	moist	+	
	oʊ	most	+	
mn	i	mnemonic (J)	—	Rare
mj	ɔ	Muriel (J)	—	mjɔr > mjur
	ʊ	Muriel	—	mjʊr > mjur
	u	music	+	
	aʊ	miaow	—	Rare
mw	ɑ	moiré	—	Foreign
	ɔ	moiré (T)	—	
n	i	neat	+	
	ɪ	knit	+	
	ɛ	neck	+	
	æ	nap	+	
	ɑ	not, narcotic	+	
	ɔ	nought	+	
	ʊ	nook	+	
	u	noose	+	
	ə	nut	+	
	ɝ	nurse	+	
	eɪ	name	+	
	aɪ	nine	+	
	aʊ	noun	+	
	ɔɪ	noise	+	
	oʊ	no	+	
nj	u	new	—	Not GA
nw	ɑ	noire (Hi)	—	Foreign
ŋ	ɑ	Ngami (J)	—	Foreign
	aɪ	Ngaio (Ho)	—	
ŋg	ɑ	Ngami (J)	—	Foreign
l	i	leaf	+	
	ɪ	lip	+	

ONSET	NUC.	KEY-WORD	MIN.	NOTES
	ɛ	let	+	
	æ	lack	+	
	ɑ	lot, lark	+	
	ɔ	law	+	
	ʊ	look	+	
	u	loom	+	
	ə	luck	+	
	ɚ	lurch	+	
	eɪ	late	+	
	aɪ	light	+	
	aʊ	lout	+	
	ɔɪ	Lloyd	+	
	oʊ	low	+	
lj	ɔ	lure (J)	—	Not GA
	ʊ	lure	—	
	u	lewd	—	
r	i	read	+	
	ɪ	rid	+	
	ɛ	red	+	
	æ	rack	+	
	ɑ	rock, Rance (J)	+	
	ɔ	raw	+	
	ʊ	rook	+	
	u	rude	+	
	ə	rum	+	
	eɪ	ray	+	
	aɪ	right	+	
	aʊ	round	+	
	ɔɪ	Roy	+	
	oʊ	road	+	
rj	u	rule (Ho)	—	Not GA
h	i	heat	+	
	ɪ	hit	+	
	ɛ	hen	+	
	æ	hat	+	
	ɑ	hot, heart	+	
	ɔ	hawk	+	
	ʊ	hood	+	
	oʊ	yolk	+	
w	i	we	+	
	ɪ	wit	+	
	ɛ	wet	+	
	æ	wag	+	
	ɑ	wallet	+	
	ɔ	walk	+	
	ʊ	would	+	
	u	woo	+	
	ə	one	+	
	ɚ	were	+	
	eɪ	wake	+	
	aɪ	wind	+	
	aʊ	wound	+	
	oʊ	woke	+	

II. Final NM Combinations

CODA	NUC.	KEY-WORD	MIN.	NOTES
zero	i	see	+	
	ɪ	happy	+	
	æ	baa (K)	+	
	ɑ	spa	+	
	ɔ	law	+	
	ʊ	value	+	
	u	two	+	
	ə	sofa	+	
	ɚ	sir	+	
	eɪ	say	+	
	aɪ	high	+	
	aʊ	how	+	
	ɔɪ	boy	+	
	oʊ	low	+	
p	i	deep	+	
	ɪ	tip	+	
	ɛ	step	+	
	æ	flap	+	
	ɑ	stop (harp)	+	
	ɔ	gaup (K)	+	

ONSET	NUC.	KEY-WORD	MIN.	NOTES
	u	hoot	+	
	ə	hut	+	
	ɝ	hurt	+	
	eɪ	hate	+	
	aɪ	high	+	
	aʊ	how	+	
	ɔɪ	hoy	+	
	oʊ	hose	+	
hj	ɑ	Hjalmar (Ho)	−	Foreign
	ʊ	Huron	+	
	u	huge	+	
hw	i	wheat	+	
	ɪ	whip	+	
	ɛ	when	+	
	æ	whack	+	
	ɑ	what	+	
	ɔ	wharf	+	
	ʊ	whoof (M)	−	Rare
	u	whoop (K)	−	Rare
	ə	whuff (M)	−	Rare
	ɝ	whirl	+	
	eɪ	whey	+	
	aɪ	why	+	
	oʊ	whoa	−	Rare
j	i	yeast	+	
	ɪ	Yiddish	+	
	ɛ	yet	+	
	æ	Yank	+	
	ɑ	yacht, yarn	+	
	ɔ	yawn	+	
	ʊ	your	+	
	u	you	+	
	ə	young	+	
	ɝ	yearn	+	
	eɪ	Yale	+	
	aɪ	yipes	−	Rare
	aʊ	yowl	+	
	ɔɪ	yoick	−	Rare

CODA	NUC.	KEY-WORD	MIN.	NOTES
	ʊ	hoop (M)	+	
	u	group	+	
	ə	cup	+	
	ɝ	chirp	+	
	eɪ	tape	+	
	aɪ	type	+	
	oʊ	hope	+	
pt	i	steeped	+	
	ɪ	tipped	+	
	ɛ	stepped	+	
	æ	wrapped	+	
	ɑ	stopped (carped)	+	
	ɔ	gauped (K)	+	
	ʊ	hooped (K)	−	ʊpt > upt
	u	grouped	+	
	ə	cupped	+	
	ɝ	chirped	+	
	eɪ	taped	+	
	aɪ	typed	+	
	oʊ	hoped	+	
pts	ɪ	scripts	+	
	ɛ	accepts	+	
	æ	adapts	+	
	ɑ	opts	+	
	ə	interrupts	+	
pf	ɪ	Zipf	−	Rare
pft	ɪ	zipfed	−	Rare
pθ	ɛ	depth	+	
pθt	ɛ	depthed (Hi)	−	Rare
pθs	ɛ	depths	−	Rare, > ps
ps	i	keeps	+	
	ɪ	tips	+	
	ɛ	steps	+	
	æ	caps	+	
	ɑ	copse (harps)	+	
	ɔ	gaups (K)	+	
	ʊ	hoops	−	ʊps > ups
	u	groups	+	

CODA	NUC.	KEY-WORD	MIN.	NOTES
	ə	cups	+	
	ɚ	chirps	+	
	eɪ	tapes	+	
	aɪ	types	+	
	oʊ	hopes	+	
pst	ɪ	eclipsed	+	
	æ	lapsed	+	
	eɪ	trapesed	+	
t	i	seat	+	
	ɪ	sit	+	
	ɛ	set	+	
	æ	sat	+	
	ɑ	pot (part)	+	
	ɔ	caught	+	
	ʊ	foot	+	
	u	boot	+	
	ə	but	+	
	ɚ	shirt	+	
	eɪ	late	+	
	aɪ	light	+	
	aʊ	about	+	
	ɔɪ	exploit	+	
	oʊ	boat	+	
tθ	ɪ	width	+	
	ɛ	breadth	+	
	ə	hundredth	+	
	eɪ	eighth	+	
tθt	ɪ	widthed	–	Rare
tθs	ɪ	widths	–	> ts
	ɛ	breadths	–	
	ə	hundredths	–	
	eɪ	eighths	–	
ts	i	seats	+	
	ɪ	sits	+	
	ɛ	sets	+	
	æ	bats	+	
	ɑ	pots (parts)	+	
	ɔ	thoughts	+	

CODA	NUC.	KEY-WORD	MIN.	NOTES
	ɑ	stock (park)	+	
	ɔ	talk	+	
	ʊ	took	+	
	u	duke	+	
	ə	duck	+	
	ɚ	work	+	
	eɪ	bake	+	
	aɪ	like	+	
	ɔɪ	hoik (J)	–	Rare
	oʊ	broke	+	
kt	i	leaked	+	
	ɪ	picked	+	
	ɛ	checked	+	
	æ	backed	+	
	ɑ	stocked (parked)	+	
	ɔ	talked	+	
	ʊ	looked	+	
	u	rebuked	+	
	ə	sucked	+	
	ɚ	worked	+	
	eɪ	baked	+	
	aɪ	liked	+	
	ɔɪ	hoiked (J)	–	Rare
	oʊ	choked	+	
kts	ɪ	convicts	+	
	ɛ	sects	+	
	æ	acts	+	
	ɑ	concocts	+	
	ə	obstructs	+	
ks	i	leaks	+	
	ɪ	six	+	
	ɛ	cheques	+	
	æ	ax	+	
	ɑ	box (barks)	+	
	ɔ	talks	+	
	ʊ	looks	+	
	u	dukes	+	
	ə	ducks	+	

CODA	NUC.	KEY-WORD	MIN.	NOTES
	ʊ	puts	+	
	u	boots	+	
	ə	cuts	+	
	ɚ	shirts	+	
	eɪ	waits	+	
	aɪ	rights	+	
	aʊ	bouts	+	
	ɔɪ	exploits	+	
	oʊ	boats	+	
tst	ɪ	blitzed	+	
tʃ	i	reach	+	
	ɪ	rich	+	
	ɛ	wretch	+	
	æ	catch	+	
	ɑ	botch (parch)	+	
	ɔ	debauch	+	
	ʊ	butch (M)	+	
	u	mooch	+	
	ə	touch	+	
	ɚ	church	+	
	eɪ	h	+	
	aʊ	couch	+	
	oʊ	coach	+	
tʃt	i	reached	+	
	ɪ	stitched	+	
	ɛ	fetched	+	
	æ	matched	+	
	ɑ	botched (parched)	+	
	ɔ	debauched	+	
	u	mooched	+	
	ə	touched	+	
	ɚ	searched	+	
	aʊ	vouched	+	
	oʊ	coached	+	
k	i	seek	+	
	ɪ	sick	+	
	ɛ	cheque	+	
	æ	lack	+	

CODA	NUC.	KEY-WORD	MIN.	NOTES
	ɚ	works	+	
	eɪ	takes	+	
	aɪ	likes	+	
	ɔɪ	hoiks (J)	−	Rare
	oʊ	hoax	+	
kst	ɪ	mixed	+	
	ɛ	text	+	
	æ	waxed	+	
	ɑ	boxed	+	
	oʊ	hoaxed	+	
ksts	ɛ	texts	+	
ksθ	ɪ	sixth	+	
	ɛ	x'th	−	Rare
ksθs	ɪ	sixths	−	> ks
	ɛ	x'ths	−	
b	i	glebe	−	Rare
	ɪ	rib	+	
	ɛ	web	+	
	æ	stab	+	
	ɑ	cob (barb)	+	
	ɔ	daub	+	
	ʊ	Danube (K)	−	ʊb > ub
	u	tube	+	
	ə	rub	+	
	ɚ	verb	+	
	eɪ	babe	+	
	aɪ	tribe	+	
	oʊ	lobe	+	
bd	ɪ	ribbed	+	
	ɛ	ebbed	+	
	æ	dabbed	+	
	ɑ	bobbed (barbed)	+	
	ɔ	daubed	+	
	ə	rubbed	+	
	ɚ	curbed	+	
	aɪ	bribed	+	
	oʊ	lobed	+	
bz	i	grebes	−	Rare

CODA	NUC.	KEY-WORD	MIN.	NOTES
	ə	cups	+	
	ɚ	chirps	+	
	eɪ	tapes	+	
	aɪ	types	+	
	oʊ	hopes	+	
pst	ɪ	eclipsed	+	
	æ	lapsed	+	
	eɪ	trapesed	+	
t	i	seat	+	
	ɪ	sit	+	
	ɛ	set	+	
	æ	sat	+	
	ɑ	pot (part)	+	
	ɔ	caught	+	
	ʊ	foot	+	
	u	boot	+	
	ə	but	+	
	ɚ	shirt	+	
	eɪ	late	+	
	aɪ	light	+	
	aʊ	about	+	
	ɔɪ	exploit	+	
	oʊ	boat	+	
tθ	ɪ	width	+	
	ɛ	breadth	+	
	ə	hundredth	+	
	eɪ	eighth	+	
tθt	ɪ	widthed	—	Rare
tθs	ɪ	widths	—	> ts
	ɛ	breadths	—	
	ə	hundredths	—	
	eɪ	eighths	—	
ts	i	seats	+	
	ɪ	sits	+	
	ɛ	sets	+	
	æ	bats	+	
	ɑ	pots (parts)	+	
	ɔ	thoughts	+	

CODA	NUC.	KEY-WORD	MIN.	NOTES
	ɑ	stock (park)	+	
	ɔ	talk	+	
	ʊ	took	+	
	u	duke	+	
	ə	duck	+	
	ɚ	work	+	
	eɪ	bake	+	
	aɪ	like	+	
	ɔɪ	hoik (J)	—	Rare
	oʊ	broke	+	
kt	i	leaked	+	
	ɪ	picked	+	
	ɛ	checked	+	
	æ	backed	+	
	ɑ	stocked (parked)	+	
	ɔ	talked	+	
	ʊ	looked	+	
	u	rebuked	+	
	ə	sucked	+	
	ɚ	worked	+	
	eɪ	baked	+	
	aɪ	liked	+	
	ɔɪ	hoiked (J)	—	Rare
	oʊ	choked	+	
kts	ɪ	convicts	+	
	ɛ	sects	+	
	æ	acts	+	
	ɑ	concocts	+	
	ə	obstructs	+	
ks	i	leaks	+	
	ɪ	six	+	
	ɛ	cheques	+	
	æ	ax	+	
	ɑ	box (barks)	+	
	ɔ	talks	+	
	ʊ	looks	+	
	u	dukes	+	
	ə	ducks	+	

CODA	NUC.	KEY-WORD	MIN.	NOTES
	ʊ	puts	+	
	u	boots	+	
	ə	cuts	+	
	ɝ	shirts	+	
	eɪ	waits	+	
	aɪ	rights	+	
	aʊ	bouts	+	
	ɔɪ	exploits	+	
	oʊ	boats	+	
tst	ɪ	blitzed	+	
tʃ	i	reach	+	
	ɪ	rich	+	
	ɛ	wretch	+	
	æ	catch	+	
	ɑ	botch (parch)	+	
	ɔ	debauch	+	
	ʊ	butch (M)	+	
	u	mooch	+	
	ə	touch	+	
	ɝ	church	+	
	eɪ	h	+	
	aʊ	couch	+	
	oʊ	coach	+	
tʃt	i	reached	+	
	ɪ	stitched	+	
	ɛ	fetched	+	
	æ	matched	+	
	ɑ	botched (parched)	+	
	ɔ	debauched	+	
	u	mooched	+	
	ə	touched	+	
	ɝ	searched	+	
	aʊ	vouched	+	
	oʊ	coached	+	
k	i	seek	+	
	ɪ	sick	+	
	ɛ	cheque	+	
	æ	lack	+	

CODA	NUC.	KEY-WORD	MIN.	NOTES
	ɝ	works	+	
	eɪ	takes	+	
	aɪ	likes	+	
	ɔɪ	hoiks (J)	−	Rare
	oʊ	hoax	+	
kst	ɪ	mixed	+	
	ɛ	text	+	
	æ	waxed	+	
	ɑ	boxed	+	
	oʊ	hoaxed	+	
ksts	ɛ	texts	+	
ksθ	ɪ	sixth	+	
	ɛ	x'th	−	Rare
ksθs	ɪ	sixths	−	> ks
	ɛ	x'ths	−	
b	i	glebe	−	Rare
	ɪ	rib	+	
	ɛ	web	+	
	æ	stab	+	
	ɑ	cob (barb)	+	
	ɔ	daub	+	
	ʊ	Danube (K)	−	ʊb > ub
	u	tube	+	
	ə	rub	+	
	ɝ	verb	+	
	eɪ	babe	+	
	aɪ	tribe	+	
	oʊ	lobe	+	
bd	ɪ	ribbed	+	
	ɛ	ebbed	+	
	æ	dabbed	+	
	ɑ	bobbed (barbed)	+	
	ɔ	daubed	+	
	ə	rubbed	+	
	ɝ	curbed	+	
	aɪ	bribed	+	
	oʊ	lobed	+	
bz	i	grebes	−	Rare

CODA	NUC.	KEY-WORD	MIN.	NOTES
	ɪ	ribs	+	
	ɛ	cobwebs	+	
	æ	dabs	+	
	ɑ	cobs (barbs)	+	
	ɔ	daubs	+	
	ʊ	Danube's	−	ʊbz > ubz
	u	tubes	+	
	ə	rubs	+	
	ɚ	verbs	+	
	eɪ	Abe's	+	
	aɪ	tribes	+	
	oʊ	lobes	+	
d	i	lead	+	
	ɪ	lid	+	
	ɛ	red	+	
	æ	bad	+	
	ɑ	cod (card)	+	
	ɔ	fraud	+	
	ʊ	could	+	
	u	mood	+	
	ə	bud	+	
	ɚ	bird	+	
	eɪ	made	+	
	aɪ	ride	+	
	aʊ	loud	+	
	ɔɪ	void	+	
	oʊ	mode	+	
dθ	ɪ	width	−	> tθ
	ɛ	breadth	−	
	ə	hundredth	−	
dθt	ɪ	widthed	−	Rare
dθs	ɪ	widths	−	> ts
	ɛ	breadths	−	
	ə	hundredths	−	
dst	ɪ	midst	−	> tst
	ɛ	dreadst (Wo)	−	
dz	i	beads	+	
	ɪ	bids	+	

CODA	NUC.	KEY-WORD	MIN.	NOTES
	aʊ	gouged	+	
g	i	league	+	
	ɪ	rig	+	
	ɛ	leg	+	
	æ	rag	+	
	ɑ	log, Prague	−	ɑg > ɔg
	ɔ	log (morgue)	+	
	u	fugue	+	
	ə	rug	+	
	ɚ	iceberg	+	
	eɪ	the Hague	+	
	oʊ	rogue	+	
gd	i	fatigued	+	
	ɪ	rigged	+	
	ɛ	begged	+	
	æ	bagged	+	
	ɑ	bogged	−	ɑgd > ɔgd
	ɔ	bogged	+	
	ə	mugged	+	
	eɪ	plagued	+	
	oʊ	rogued (K)	−	Rare
gz	i	leagues	+	
	ɪ	rigs	+	
	ɛ	legs	+	
	æ	bags	+	
	ɑ	bogs	−	ɑgz > ɔgz
	ɔ	bogs (morgues)	+	
	u	fugues	+	
	ə	rugs	+	
	ɚ	ergs	+	
	eɪ	Craig's	+	
	oʊ	rogues	+	
f	i	leaf	+	
	ɪ	stiff	+	
	ɛ	chef	+	
	æ	laugh	+	
	ɑ	laugh (cough: ɒ)	−	ɑf > æf, ɒf > ɔf
	ɔ	cough	+	

CODA	NUC.	KEY-WORD	MIN.	NOTES
	ɛ	beds	+	
	æ	fads	+	
	ɑ	cods (cards)	+	
	ɔ	applauds	+	
	ʊ	woods	+	
	u	moods	+	
	ə	buds	+	
	ɚ	birds	+	
	eɪ	trades	+	
	aɪ	rides	+	
	aʊ	crowds	+	
	ɔɪ	Lloyd's	+	
	oʊ	modes	+	
dzd	æ	adzed	–	Rare
dʒ	i	siege	+	
	ɪ	ridge	+	
	ɛ	hedge	+	
	æ	badge	+	
	ɑ	lodge (large)	+	
	(ɔ	forge: r-less)	–	Not GA
	u	huge	+	
	ə	fudge	+	
	ɚ	merge	+	
	eɪ	rage	+	
	aɪ	oblige	+	
	aʊ	gouge	+	
	oʊ	doge	–	Rare
dʒd	i	besieged	+	
	ɪ	ridged	+	
	ɛ	edged	+	
	æ	badged	+	
	ɑ	dodged (enlarged)	+	
	(ɔ	forged: r-less)	–	Not GA
	u	rouged	+	
	ə	judged	+	
	ɚ	merged	+	
	eɪ	raged	+	
	aɪ	obliged	+	

CODA	NUC.	KEY-WORD	MIN.	NOTES
	ʊ	whoof (M)	–	Rare, ʊf > uf
	u	roof	+	
	ə	rough	+	
	ɚ	surf	+	
	eɪ	safe	+	
	aɪ	life	+	
	ɔɪ	coif	–	Rare
	oʊ	loaf	+	
ft	i	beefed	+	
	ɪ	lift	+	
	ɛ	left	+	
	æ	draft	+	
	ɑ	draft (coughed: ɒ)	–	ɑft > æft, ɒft > ɔft
	ɔ	coughed	+	
	ʊ	hoofed (K)	–	ʊft > uft
	u	goofed	+	
	ə	stuffed	+	
	ɚ	surfed	+	
	eɪ	chafed	+	
	aɪ	knifed	+	
	ɔɪ	coifed	–	Rare
	oʊ	loafed	+	
fts	ɪ	lifts	+	
	ɛ	thefts	+	
	æ	rafts	+	
	ɑ	rafts	–	ɑfts > æfts
	ɔ	crofts	+	
	ə	tufts	+	
fθ	ɪ	fifth	+	
	ɛ	f'th (Wh)	+	
fθs	ɪ	fifths	–	> fs
	ɛ	f'ths	–	
fs	i	reefs	+	
	ɪ	cliffs	+	
	ɛ	chefs	+	
	æ	laughs	+	
	ɑ	laughs (coughs: ɒ)	–	ɑfs > æfs, ɒfs > ɔfs
	ɔ	coughs	+	

CODA	NUC.	KEY-WORD	MIN.	NOTES
	ʊ	hoofs (M)	−	ʊfs > ufs
	u	roofs	+	
	ə	stuffs	+	
	ɚ	surfs	+	
	eɪ	chafes	+	
	aɪ	wife's	+	
	ɔɪ	coifs	−	Rare
	oʊ	oafs	+	
θ	i	wreath	+	
	ɪ	smith	+	
	ɛ	death	+	
	æ	path	+	
	ɑ	path (cloth: ɒ)	−	ɑθ > æθ, ɒθ > ɔθ
	ɔ	cloth (forth)	+	
	u	uncouth	+	
	ə	doth	−	Rare
	ɚ	earth	+	
	eɪ	faith	+	
	aɪ	myth (Wo, J)	−	Rare
	aʊ	mouth	+	
	oʊ	oath	+	
θt	ɛ	breathed	+	
	æ	bathed	+	
	ɑ	bathed	−	ɑθt > æθt
	u	toothed	−	Rare
	ɚ	berthed	+	
θs	i	Keith's	+	
	ɪ	smiths	+	
	ɛ	deaths	+	
	æ	baths (verb)	+	
	ɑ	baths (cloths: ɒ)	−	ɑθs > æθs, ɒθs > ɔθs
	ɔ	cloths (fourths)	+	
	u	youth's	+	
	ɚ	earths	+	
	eɪ	faiths	+	
	aɪ	myths (Wo, J)	−	Rare
	aʊ	mouth's	−	Rare

CODA	NUC.	KEY-WORD	MIN.	NOTES
	ɚ	first	+	
	eɪ	taste	+	
	aɪ	iced	+	
	aʊ	oust	+	
	ɔɪ	hoist	+	
	oʊ	host	+	
sts	i	beasts	+	
	ɪ	lists	+	
	ɛ	rests	+	
	æ	lasts	+	
	ɑ	lasts (costs: ɒ)	−	ɑsts > æsts, ɒsts > ɔsts
	ɔ	exhausts	+	
	u	boosts	+	
	ə	dusts	+	
	ɚ	bursts	+	
	eɪ	tastes	+	
	aɪ	Christs's	+	
	aʊ	ousts	+	
	ɔɪ	hoists	+	
	oʊ	hosts	+	
sk	ɪ	risk	+	
	ɛ	desk	+	
	æ	ask	+	
	ɑ	ask, mosque	−	ɑsk > æsk or ɔsk
	ɔ	mosque	+	
	ʊ	brusque	−	Rare
	ə	rusk	+	
skt	ɪ	risked	+	
	ɛ	burlesqued	+	
	æ	asked	+	
	ɑ	asked	−	ɑskt > æskt
	ə	tusked	+	
sks	ɪ	risks	+	
	ɛ	desks	+	
	æ	asks	+	
	ɑ	asks, mosques	−	ɑsks > æsks or ɔsks
	ɔ	mosques	+	

CODA	NUC.	KEY-WORD	MIN.	NOTES
	ou	growths	+	
s	i	lease	+	
	ɪ	hiss	+	
	ɛ	less	+	
	æ	gas	+	
	ɑ	boss (K), glass	+	
	ɔ	sauce	+	
	ʊ	puss (M)	+	
	u	moose	+	
	ə	fuss	+	
	ɚ	curse	+	
	eɪ	base	+	
	aɪ	ice	+	
	aʊ	house	+	
	ɔɪ	voice	+	
	ou	close	+	
sp	ɪ	lisp	+	
	æ	clasp	+	
	ɑ	clasp, wasp	—	ɑsp > æsp or ɔsp
	ɔ	wasp	+	
	ə	cusp	+	
spt	ɪ	lisped	+	
	æ	clasped	+	
	ɑ	clasped	—	ɑspt > æspt
	ə	cusped	+	
sps	ɪ	lisps	+	
	æ	clasps	+	
	ɑ	clasps, wasps	—	ɑsps > æsps or ɔsps
	ɔ	wasps	+	
	ə	cusps	+	
st	i	least	+	
	ɪ	list	+	
	ɛ	best	+	
	æ	cast	+	
	ɑ	cast (cost: ɒ)	—	ɑst > æst, ɒst > ɔst
	ɔ	exhaust	+	
	u	boost	+	
	ə	dust	+	

CODA	NUC.	KEY-WORD	MIN.	NOTES
	ə	tusks	+	
ʃ	i	leash	+	
	ɪ	fish	+	
	ɛ	fresh	+	
	æ	rash	+	
	ɑ	wash (harsh)	+	
	ɔ	wash	—	ɔʃ > ɑʃ
	ʊ	push	+	
	u	ruche (Ho)	—	Rare
	ə	rush	+	
	eɪ	créche	—	Rare
	ou	gauche	—	Rare
ʃt	i	leashed	+	
	ɪ	fished	+	
	ɛ	meshed	+	
	æ	smashed	+	
	ɑ	washed (harshed)	+	
	ɔ	washed	—	ɔʃt > ɑʃt
	ʊ	pushed	+	
	ə	rushed	+	
v	i	grieve	+	
	ɪ	live	+	
	æ	halve	+	
	ɑ	of, halve	+	
	u	move	+	
	ə	love	+	
	ɚ	curve	+	
	eɪ	save	+	
	aɪ	drive	+	
	ou	stove	+	
vd	i	believed	+	
	ɪ	lived	+	
	æ	halved	+	
	ɑ	halved (carved)	—	ɑvd > ævd
	u	moved	+	
	ə	loved	+	
	ɚ	curved	+	
	eɪ	saved	+	

CODA	NUC.	KEY-WORD	MIN.	NOTES	CODA	NUC.	KEY-WORD	MIN.	NOTES
	aɪ	contrived	+			ɑ	garage	+	
	oʊ	roved	−	Rare		u	rouge	+	
vdz	i	(the) bereaved's	−	Rare		eɪ	beige	+	
vz	i	eaves	+			oʊ	loge (Ho)	+	
	ɪ	lives	+		ʒd	ɑ	garaged	−	Rare, > dʒd
	æ	halves	+			u	roughed	−	
	ɑ	halves (carves)	−	ɑvz > ævz		eɪ	beiged	−	
	u	moves	+		m	i	steam	+	
	ə	loves	+			ɪ	swim	+	
	ɚ	curves	+			ɛ	stem	+	
	eɪ	saves	+			æ	ham	+	
	aɪ	lives	+			ɑ	bomb, balm	+	
	aʊ	loaves	+			ɔ	haulm (warm)	−	Rare
ð	i	seethe	+			ʊ	room	−	ʊm > um
	ɪ	with	+			u	loom	+	
	ɛ	eth (M)	−	Rare		ə	drum	+	
	u	smooth	+			ɚ	firm	+	
	eɪ	bathe	+			eɪ	tame	+	
	aɪ	lithe	+			aɪ	time	+	
	aʊ	mouth (verb)	+			oʊ	loam	+	
	oʊ	loathe	+		mp	ɪ	limp	+	
ðd	i	breathed	+			ɛ	hemp	+	
	u	smoothed	+			æ	hamp	+	
	eɪ	bathed	+			ɑ	romp	+	
	aɪ	tithed (J)	−	Rare		ɔ	swamp	+	
	aʊ	mouthed	+			ə	hump	+	
	oʊ	loathed	+		mpt	ɪ	limped	+	
ðz	i	breathes	+			ɛ	tempt	+	
	ɛ	eths	−	Rare		æ	camped	+	
	æ	baths (M)	+			ɑ	prompt	+	
	(ɑ	moths: ɒ)	−	ɒðz > ɔðz		ɔ	swamped	+	
	ɔ	moths	+			ə	bumped	+	
	u	smooths	+		mpts	ɛ	tempts	+	
	eɪ	bathes	+			ɑ	prompts	+	
	aɪ	tithes	+		mpf	ɪ	nymph	−	> mf
	aʊ	mouths (verb)	+			æ	Banff	−	
	oʊ	loathes	+			ʊ	oomph	−	
ðm	ɪ	rhythm	−	> ðəm		u	oomph	−	

CODA	NUC.	KEY-WORD	MIN.	NOTES
ðmd	ɪ	rhythmed	–	> ðəmd
ðmz	ɪ	rhythms	–	> ðəmz
z	i	breeze	+	
	ɪ	his	+	
	ɛ	says	+	
	æ	has	+	
	ɑ	Ros, vase	+	
	ɔ	cause	+	
	u	lose	+	
	ə	buzz	+	
	ɚ	hers	+	
	eɪ	raise	+	
	aɪ	rise	+	
	aʊ	house (verb)	+	
	ɔɪ	noise	+	
	oʊ	rose	+	
zd	i	breezed	+	
	ɪ	fizzed	+	
	ɛ	fezzed	+	
	æ	jazzed	+	
	ɔ	caused	+	
	u	fused	+	
	ə	buzzed	+	
	ɚ	furzed (M)	–	Rare
	eɪ	raised	+	
	aɪ	surmised	+	
	aʊ	housed	+	
	ɔɪ	poised	+	
	oʊ	supposed	+	
zm	ɪ	socialism	–	> zəm
	æ	chasm	–	
	ɑ	microcosm	–	
zmd	æ	chasmed	–	> zəmd
zmz	ɪ	isms	–	> zəmz
	æ	chasms	–	
	ɑ	microcosms	–	
ʒ	i	prestige	+	
	ɛ	manège (K)	–	Foreign

CODA	NUC.	KEY-WORD	MIN.	NOTES
	ə	humph (Wh)	–	
mpft	ʊ	oomphed	–	> mft
	u	oomphed	–	
	ə	triumphed	–	
mpfs	ɪ	nymphs	–	> mfs
	ʊ	oomphs	–	
	u	oomphs	–	
	ə	triumphs	–	
mpfst	ə	triumphst	–	Rare
mps	ɪ	glimpse	+	
	æ	lamps	+	
	ɑ	romps	+	
	ɔ	swamps	+	
	ə	bumps	+	
mpst	ɪ	glimpsed	+	
mt	ɪ	limped	–	> mpt
	ɛ	tempt	–	
	æ	camped	–	
	ɑ	prompt	–	
	ə	bumped	–	
mts	ɛ	tempts	–	> mpts
	ɑ	prompts	–	
mb	æ	iamb	–	Rare, > m
	ɑ	rhomb	–	
md	i	dreamed	+	
	ɪ	dimmed	+	
	ɛ	stemmed	+	
	æ	crammed	+	
	ɑ	bombed (harmed)	+	
	(ɔ	formed: r-less)	–	Not GA
	u	loomed	+	
	ə	drummed	+	
	ɚ	affirmed	+	
	eɪ	famed	+	
	aɪ	timed	+	
	oʊ	roamed	+	
mf	ɪ	nymph	+	
	æ	Banff	–	Rare

CODA	NUC.	KEY-WORD	MIN.	NOTES
	ʊ	oomph	−	Rare
	u	oomph	−	Rare
	ə	triumph	+	
mft	ʊ	oomphed	−	Rare
	u	oomphed	−	Rare
	ə	triumphed	+	
mfs	ɪ	nymphs	+	
	ʊ	oomphs	−	Rare
	u	oomphs	−	Rare
	ə	triumphs	+	
mfst	ə	triumphst	−	Rare
mθ	ɛ	m'th	−	Rare
	(ɔ	warmth: r-less)	−	Not GA
(mθs	ɔ	warmths: r-less)	−	Not GA
ms	ɪ	glimpse	−	> mps
	æ	lamps	−	
	ɑ	romps	−	
	ɔ	swamps	−	
	ə	bumps	−	
mst	ɪ	glimpsed	−	> mpst
mz	i	dreams	+	
	ɪ	swims	+	
	ɛ	stems	+	
	æ	crams	+	
	ɑ	bombs, balms	+	
	ɔ	haulms (forms)	−	Rare
	ʊ	rooms	−	ʊmz > umz
	u	looms	+	
	ə	comes	+	
	ɚ	firms	+	
	eɪ	games	+	
	aɪ	times	+	
	oʊ	roams	+	
n	i	seen	+	
	ɪ	sin	+	
	ɛ	pen	+	
	æ	pan	+	
	ɑ	don (barn)	+	

CODA	NUC.	KEY-WORD	MIN.	NOTES
	aɪ	pints	+	
	aʊ	counts	+	
	ɔɪ	joints	+	
	oʊ	don't's	+	
ntst	ɪ	chintzed, rinced	−	> nst
	ɛ	fenced	−	
	æ	glanced	−	
	ɑ	sconced, glanced	−	
	eɪ	against	−	
	aʊ	pounced	−	
ntʃ	ɪ	inch	+	
	ɛ	bench	+	
	æ	branch	+	
	ɑ	branch	−	ɑntʃ > æntʃ
	ɔ	staunch	+	
	ʊ	Munch	−	Rare
	ə	lunch	+	
ntʃt	ɪ	inched	+	
	ɛ	wrenched	+	
	æ	branched	+	
	ɑ	branched	−	ɑntʃt > æntʃt
	ɔ	launched	+	
	ə	lunched	+	
nd	i	fiend	+	
	ɪ	wind	+	
	ɛ	bend	+	
	æ	band	+	
	ɑ	bond, demand	+	
	ɔ	dawned	+	
	u	wound	+	
	ə	fund	+	
	ɚ	burned	+	
	eɪ	reigned	+	
	aɪ	lined	+	
	aʊ	bound	+	
	ɔɪ	joined	+	
	oʊ	toned	+	
ndθ	ɔ	thousandth	−	Rare, > nθ

CODA	NUC.	KEY-WORD	MIN.	NOTES
	ɔ	dawn	+	
	u	moon	+	
	ə	done	+	
	ɚ	burn	+	
	eɪ	lane	+	
	aɪ	line	+	
	aʊ	down	+	
	ɔɪ	join	+	
	oʊ	bone	+	
nt	ɪ	hint	+	
	ɛ	bent	+	
	æ	cant	+	
	ɑ	font, can't	+	
	ɔ	jaunt	+	
	ə	stunt	+	
	ɚ	burnt	+	
	eɪ	paint	+	
	aɪ	pint	+	
	aʊ	count	+	
	ɔɪ	point	+	
	oʊ	don't	+	
ntθ	i	thirteenth	–	> nθ
	ɪ	plinth	–	
	ɛ	tenth	–	
	ə	month	–	
	aɪ	ninth	–	
ntθs	i	thirteenths	–	> nθs
	ɪ	plinths	–	
	ɛ	tenths	–	
	ə	months	–	
	aɪ	ninths	–	
nts	ɪ	hints	+	
	ɛ	tents	+	
	æ	ants	+	
	ɑ	fonts, ants	+	
	ɔ	jaunts	+	
	ə	stunts	+	
	eɪ	taints	+	
ndθs	ə	thousandths	–	Rare, > ns ?
ndz	i	fiends	+	
	ɪ	winds	+	
	ɛ	friends	+	
	æ	lands	+	
	ɑ	bonds, demands	+	
	u	wounds	+	
	ə	funds	+	
	aɪ	kinds	+	
	aʊ	sounds	+	
ndʒ	ɪ	singe	+	
	ɛ	revenge	+	
	æ	flange	+	
	ɑ	sponge	+	
	ə	plunge	+	
	eɪ	change	+	
	aʊ	lounge	+	
ndʒd	ɪ	hinged	+	
	ɛ	revenged	+	
	æ	flanged	+	
	ɑ	sponged	+	
	ə	plunged	+	
	eɪ	changed	+	
	aʊ	lounged	+	
nf	æ	Banff (J)	–	Rare
nθ	i	thirteenth	+	
	ɪ	plinth	+	
	ɛ	tenth	+	
	ə	month	+	
	aɪ	ninth	+	
nθs	i	thirteenths	–	> ns, nts
	ɪ	plinths	–	
	ɛ	tenths	–	
	ə	months	–	
	aɪ	ninths	–	
ns	ɪ	prince	+	
	ɛ	sense	+	
	æ	dance	+	

CODA	NUC.	KEY-WORD	MIN.	NOTES
	ɑ	sconce, dance	+	
	ə	dunce	+	
	aɪ	ninths	+	
	aʊ	ounce	+	
nst	ɪ	rinced	+	
	ɛ	fenced	+	
	æ	glanced	+	
	ɑ	sconced, glanced	+	
	ə	once	−	ənst > əns
	eɪ	against	+	
	aʊ	pounced	+	
nsk	ɪ	Minsk	−	Foreign
nʃ	ɪ	inch	−	> ntʃ
	ɛ	bench	−	
	æ	branch	−	
	ɑ	branch	−	
	ɔ	staunch	−	
	ʊ	Munch	−	
	ə	lunch	−	
nʃt	ɪ	inched	−	> ntʃt
	ɛ	wrenched	−	
	æ	branched	−	
	ɑ	branched	−	
	ɔ	launched	−	
	ə	lunched	−	
nz	i	scenes	+	
	ɪ	sins	+	
	ɛ	pens	+	
	æ	pans	+	
	ɑ	dons (barns)	+	
	ɔ	dawns	+	
	u	moons	+	
	ə	buns	+	
	ɚ	burns	+	
	eɪ	lanes	+	
	aɪ	lines	+	
	aʊ	downs	+	
	ɔɪ	joins	+	

CODA	NUC.	KEY-WORD	MIN.	NOTES
	ə	flunked	+	
ŋkts	ɪ	instincts	+	
ŋkθ	ɛ	length	+	
ŋkθt	ɛ	lengthed	−	Rare
ŋkθs	ɛ	lengths	−	> ŋks
ŋks	ɪ	lynx	+	
	ɛ	Schenck's	+	
	æ	ranks	+	
	ɑ	conks	−	ɑŋks > ɔŋks
	ɔ	conks	+	
	ə	trunks	+	
ŋkst	ɪ	jinxed	+	
	ɑ	wrongst	−	Rare
	ɔ	wrongst	−	Rare
	ə	amongst	+	
ŋd	ɪ	ringed	+	
	æ	banged	+	
	ɑ	belonged	−	ɑŋd > ɔŋd
	ɔ	belonged	+	
	ə	lunged	+	
ŋθ	ɛ	length	−	> ŋkθ
ŋθt	ɛ	lengthed	−	Rare
ŋθs	ɛ	lengths	−	> ŋks
ŋs	ɪ	lynx	−	> ŋks
	ɛ	Schenck's	−	
	æ	ranks	−	
	ɑ	conks	−	
	ɔ	conks	−	
	ə	bunks	−	
ŋst	ɪ	jinxed	−	> ŋkst
	ɑ	wrongst (Wo)	−	
	ɔ	wrongst (Wo)	−	
	ə	amongst	−	
ŋz	ɪ	wings	+	
	æ	bangs	+	
	ɑ	belongs	−	ɑŋz > ɔŋz
	ɔ	belongs	+	
	ə	lungs	+	

CODA	NUC.	KEY-WORD	MIN.	NOTES
	oʊ	bones	+	
nzd	ɛ	cleansed	+	
	ɑ	bronzed	+	
nʒ	ɪ	singe	−	> ndʒ
	ɛ	revenge	−	
	æ	flange	−	
	ɑ	sponge	−	
	ə	plunge	−	
	eɪ	arrange	−	
	aʊ	scrounge	−	
nʒd	ɪ	hinged	−	> ndʒd
	ɛ	revenged	−	
	æ	flanged	−	
	ɑ	sponged	−	
	ə	plunged	−	
	eɪ	arranged	−	
	aʊ	scrounged	−	
ŋ	ɪ	sing	+	
	æ	bang	+	
	ɑ	long	−	ɑŋ > ɔŋ
	ɔ	long	+	
	ə	dung	+	
ŋt	ɪ	linked	−	> ŋkt
	æ	ranked	−	
	ɑ	conked	−	
	ɔ	conked	−	
	ə	flunked	−	
ŋts	ɪ	instincts	−	> ŋkts
ŋk	ɪ	link	+	
	ɛ	Schenck	+	
	æ	rank	+	
	ɑ	conk	−	ɑŋk > ɔŋk
	ɔ	conk	+	
	ə	bunk	+	
ŋkt	ɪ	linked	+	
	æ	ranked	+	
	ɑ	conked	−	ɑŋkt > ɔŋkt
	ɔ	conked	+	

CODA	NUC.	KEY-WORD	MIN.	NOTES
l	i	steel	+	
	ɪ	still	+	
	ɛ	bell	+	
	æ	pal	+	
	ɑ	doll, morale	+	
	ɔ	call	+	
	ʊ	pull	+	
	u	pool	+	
	ə	dull	+	
	ɚ	curl	+	
	eɪ	pale	+	
	aɪ	pile	+	
	aʊ	foul	+	
	ɔɪ	foil	+	
	oʊ	roll	+	
lp	ɛ	help	+	
	æ	scalp	+	
	u	poulpe	−	Rare
	ə	pulp	+	
	oʊ	holp	−	Rare
lpt	ɛ	helped	+	
	æ	scalped	+	
	ə	gulped	+	
lps	ɛ	helps	+	
	æ	scalps	+	
	u	poulpes	−	Rare
	ə	gulps	+	
lt	ɪ	built	+	
	ɛ	felt	+	
	æ	shalt	−	Rare
	(ɑ	malt: ɒ)	−	ɒlt > ɔlt
	ɔ	malt	+	
	ə	cult	+	
	ɔɪ	spoilt	+	
	oʊ	bolt	+	
lts	ɪ	kilts	+	
	ɛ	melts	+	
	(ɑ	waltz: ɒ)	−	ɒlts > ɔlts

CODA	NUC.	KEY-WORD	MIN.	NOTES
	ɔ	waltz	+	
	ə	cults	+	
	oʊ	bolts	+	
ltst	(ɑ	waltzed: ɒ)	−	> lst
	ɔ	waltzed	−	
ltʃ	ɪ	filch	+	
	ɛ	belch	+	
	ɔ	Balch (Wo)	−	Rare
	ə	mulch	+	
ltʃt	ɪ	filched	+	
	ɛ	belched	+	
	ə	mulched	+	
lk	ɪ	milk	+	
	ɛ	elk	+	
	æ	talc	+	
	ə	bulk	+	
lkt	ɪ	milked	+	
	æ	calqued	+	
	ə	mulct	+	
lkts	ə	mulcts	−	Rare, > lks
lks	ɪ	milks	+	
	ɛ	elks	+	
	æ	calx	+	
	ə	bulks	+	
lb	æ	alb	+	
	ə	bulb	+	
lbd	æ	albed	−	Rare
	ə	bulbed	+	
lbz	æ	albs	−	Rare
	ə	bulbs	+	
ld	i	field	+	
	ɪ	filled	+	
	ɛ	felled	+	
	æ	palled	+	
	ɑ	dolled (snarled)	+	
	ɔ	bald	+	
	ʊ	pulled	+	
	u	pooled	+	

CODA	NUC.	KEY-WORD	MIN.	NOTES
	ə	gulfs	+	
lθ	ɪ	filth	+	
	ɛ	health	+	
	u	coolth (M)	−	Rare
lθt	ɛ	healthed	−	Rare
lθs	ɛ	wealths	−	> ls
ls	ɪ	grilse (M)	−	Rare
	ɛ	else	+	
	æ	salse (M)	−	Rare
	(ɑ	false: ɒ)	−	ɒls > ɔls
	ɔ	false	+	
	ə	pulse	+	
lst	ə	pulsed	+	
	aɪ	whilst	−	Rare
lʃ	ɛ	Welsh	+	
	ɔ	Walsh	+	
lʃt	ɛ	welshed	−	Rare, > ltʃt
lv	ɛ	twelve	+	
	æ	valve	+	
	ɑ	solve	+	
lvd	ɛ	delved	+	
	æ	salved	+	
	ɑ	solved	+	
lvz	ɛ	elves	+	
	æ	valves	+	
	ɑ	solves	+	
	ʊ	wolves	+	
lz	i	steals	+	
	ɪ	pills	+	
	ɛ	bells	+	
	æ	pals	+	
	ɑ	dolls (snarls)	+	
	ɔ	stalls	+	
	ʊ	pulls	+	
	u	pools	+	
	ə	gulls	+	
	ɚ	curls	+	
	eɪ	tails	+	

CODA	NUC.	KEY-WORD	MIN.	NOTES
	ə	mulled	+	
	ɚ	curled	+	
	eɪ	mailed	+	
	aɪ	piled	+	
	aʊ	fouled	+	
	ɔɪ	soiled	+	
	oʊ	fold	+	
ldz	i	fields	+	
	ɪ	builds	+	
	ɛ	welds	+	
	ɔ	scalds	+	
	ɚ	worlds	+	
	aɪ	wilds	+	
	oʊ	folds	+	
ldʒ	ɪ	bilge	+	
	ə	bulge	+	
ldʒd	ɪ	bilged	+	
	ə	bulged	+	
lf	ɪ	sylph	+	
	ɛ	self	+	
	æ	Ralph	+	
	ɑ	golf	+	
	ɔ	golf	−	ɔlf > ɑlf
	ʊ	wolf	+	
	ə	gulf	+	
lft	ɛ	Delft	+	
	ɑ	golfed	+	
	ɔ	golfed	−	ɔlft > ɑlft
	ʊ	wolfed	+	
	ə	engulfed	+	
lfθ	ɛ	twelfth	+	
lfθs	ɛ	twelfths	−	> lfs
lfs	ɪ	sylphs	+	
	ɛ	elf's	+	
	æ	Ralph's	+	
	ɑ	golfs	+	
	ɔ	golfs	−	ɔlfs > ɑlfs
	ʊ	wolf's	+	

CODA	NUC.	KEY-WORD	MIN.	NOTES
	aɪ	piles	+	
	aʊ	howls	+	
	ɔɪ	soils	+	
	oʊ	rolls	+	
lʒ	ɪ	bilge (?)	−	> ldʒ
	ə	bulge (?)	−	
lʒd	ɪ	bilged (?)	−	> ldʒd
	ə	bulged (?)	−	
lm	ɪ	film	+	
	ɛ	realm	+	
	ə	culm	−	Rare
lmd	ɪ	filmed	+	
	ɛ	overwhelmed	+	
lmz	ɪ	films	+	
	ɛ	realms	+	
	ə	culms	−	Rare
ln	ɪ	kiln	−	Rare, > l
	oʊ	Colne, swollen	−	
lnd	ɪ	kilned	−	Rare, > ld
lnz	ɪ	kilns	−	Rare, > lz
r	i	beer	−	ir > ɪr
	ɪ	beer	+	
	ɛ	bear	+	
	æ	bear	−	ær > ɛr
	ɑ	bar	+	
	ɔ	for	+	
	ʊ	poor	+	
	u	poor	−	ur > ʊr
	eɪ	air, layer	−	eɪr > ɛr, eɪɚ
	aɪ	fire	−	aɪr > aɪɚ
	aʊ	power	−	aʊr > aʊɚ
	oʊ	four, lower	−	oʊr > ɔr, oʊɚ
rp	ɑ	carp	+	
	ɔ	warp	+	
rpt	ɑ	carped	+	
	ɔ	warped	+	
rps	ɑ	harps	+	
	ɔ	corpse	+	

CODA	NUC.	KEY-WORD	MIN.	NOTES
rt	ɑ	heart	+	
	ɔ	tort	+	
	oʊ	court	−	oʊrt > ɔrt
rts	ɑ	hearts	+	
	ɔ	quartz	+	
	oʊ	courts	−	oʊrts > ɔrts
rtst	ɔ	quartzed	−	Rare
rtʃ	ɑ	arch	+	
	ɔ	scorch	+	
	oʊ	porch	−	oʊrtʃ > ɔrtʃ
rtʃt	ɑ	parched	+	
	ɔ	scorched	+	
rk	ɑ	park	+	
	ɔ	fork	+	
	u	Bourke (Wo)	−	Rare
	oʊ	pork	−	oʊrk > ɔrk
rkt	ɑ	parked	+	
	ɔ	forked	+	
rks	ɑ	parks	+	
	ɔ	forks	+	
	u	Bourke's (Wo)	−	Rare
	oʊ	porks	−	oʊrks > ɔrks
rb	ɑ	garb	+	
	ɔ	orb	+	
rbd	ɑ	barbed	+	
	ɔ	orbed	+	
rbz	ɑ	garbs	+	
	ɔ	orbs	+	
rd	i	steered	−	ird > ɪrd
	ɪ	steered	+	
	ɛ	cared	+	
	æ	cared	−	ærd > ɛrd
	ɑ	guard	+	
	ɔ	cord	+	
	ʊ	toured	+	
	u	toured	−	urd > ʊrd
	eɪ	aired, layered	−	eɪrd > ɛrd, eɪɚd
	aɪ	fired	−	aɪrd > aɪɚd

CODA	NUC.	KEY-WORD	MIN.	NOTES
	ɑ	farced	+	
	ɔ	endorsed	+	
	oʊ	forced	−	oʊrst > ɔrst
rsk	ɔ	torsk (Wo)	−	Foreign
rʃ	ɑ	harsh	+	
rʃt	ɑ	harshed (M)	+	
rv	ɑ	carve	+	
rvd	ɑ	carved	+	
rvz	ɑ	carves	+	
	ɔ	wharves	+	
rz	i	fears	−	irz > ɪrz
	ɪ	fears	+	
	ɛ	cares	+	
	æ	cares	−	ærz > ɛrz
	ɑ	cars	+	
	ɔ	wars	+	
	ʊ	tours	+	
	u	tours	−	urz > ʊrz
	eɪ	airs, layers	−	eɪrz > ɛrz, eɪɚz
	aɪ	fires	−	aɪrz > aɪɚz
	aʊ	powers	−	aʊrz > aʊɚz
	oʊ	doors, lowers	−	oʊrz > ɔrz, oʊɚz
rm	ɑ	farm	+	
	ɔ	dorm	+	
rmpθ	ɔ	warmth	−	> rmθ
rmpθs	ɔ	warmths	−	> rmθs
rmd	ɑ	farmed	+	
	ɔ	formed	+	
rmθ	ɔ	warmth	+	
rmθs	ɔ	warmths	+	
rmz	ɑ	farms	+	
	ɔ	forms	+	
rn	ɛ	cairn	+	
	æ	cairn	−	ærn > ɛrn
	ɑ	barn	+	
	ɔ	corn	+	
	eɪ	cairn	−	eɪrn > ɛrn
	oʊ	mourn	−	oʊrn > ɔrn

CODA	NUC.	KEY-WORD	MIN.	NOTES
	aʊ	powered	−	aʊrd > aʊɚd
	oʊ	board, lowered	−	oʊrd > ɔrd, oʊɚd
rdz	i	beards	−	irdz > ɪrdz
	ɪ	beards	+	
	ɛ	lairds	+	
	æ	lairds	−	ærdz > ɛrdz
	ɑ	cards	+	
	ɔ	cords	+	
	oʊ	boards	−	oʊrdz > ɔrdz
rdʒ	ɑ	charge	+	
	ɔ	forge	+	
rdʒd	ɑ	charged	+	
	ɔ	forged	+	
rg	ɑ	Sarg (M)	−	Rare
	ɔ	morgue	−	Rare
rgz	ɔ	morgues	−	Rare
rf	ɑ	scarf	+	
	ɔ	morph	+	
rft	ɑ	scarfed	+	
	ɔ	dwarfed	+	
rfs	ɑ	scarfs	+	
	ɔ	morphs	+	
rθ	ɑ	hearth	+	
	ɔ	Orth (M)	+	
	oʊ	fourth	−	oʊrθ > ɔrθ
rθs	ɑ	hearths	+	
	ɔ	Orth's	+	
	oʊ	fourths	−	oʊrθs > ɔrθs
rs	i	pierce	−	irs > ɪrs
	ɪ	pierce	+	
	ɛ	scarce	+	
	æ	scarce	−	ærs > ɛrs
	ɑ	farce	+	
	ɔ	Morse	+	
	ʊ	bourse	+	
	oʊ	force	−	oʊrs > ɔrs
rst	i	pierced	−	irst > ɪrst
	ɪ	pierced	+	

CODA	NUC.	KEY-WORD	MIN.	NOTES
rnt	ɑ	aren't	+	
rnd	ɛ	cairned	+	
	æ	cairned	−	ærnd > ɛrnd
	ɑ	darned	+	
	ɔ	horned	+	
	eɪ	cairned	−	eɪrnd > ɛrnd
	oʊ	mourned	−	oʊrnd > ɔrnd
rnz	ɛ	cairns	+	
	æ	cairns	−	ærnz > ɛrnz
	ɑ	barns	+	
	ɔ	horns	+	
	eɪ	cairns	−	eɪrnz > ɛrnz
	oʊ	mourns	−	oʊrnz > ɔrnz
rl	ɑ	snarl	+	
	ɔ	orle	−	Rare
rld	ɑ	snarled	+	
rlz	ɑ	snarls	+	
	ɔ	orles	−	Rare

THE COLLEGE OF SPEECH THERAPISTS

The College of Speech Therapists is holding a Conference in Birmingham, England, from 24th to 28th July, 1961.

Abnormalities of speech and language will be presented as systemic disorders under the guidance of the concepts of Signs, Signals and Symbols.

Further information may be obtained from The Conference Secretary, 16 York Road, Birmingham 16, England.

Professor J. L. Pauwels, Editor of *Leuvense Bijdragen*, wishes to obtain *Language and Speech*, Vol. 1, Part 4 (Oct.-Dec. 1958) to complete his collection. Offers should be addressed to him at Naamse Vest 40, Leuven, Belgium.

ADDRESSES OF CONTRIBUTORS TO VOL. 4, PART 1

BRUCE, Dr. D. J., Department of Psychology, The University, Reading, England.

GOLDMAN-EISLER, Dr. Frieda, Department of Phonetics, University College, Gower Street, London W.C.1, England.

SIVERTSEN, Dr. Eva, Norges Lærerhøgskole, Trondheim, Norway.

SOME LINGUISTIC FEATURES OF SPEECH FROM APHASIC PATIENTS*

Samuel Fillenbaum and Lyle V. Jones
University of North Carolina

and

Joseph M. Wepman
University of Chicago

The free speech of each of twelve adult aphasic patients was examined with reference particularly to (1) the distribution of words according to grammatical function, (2) sequential dependencies in form-class usage, and (3) stereotypy in vocabulary. The majority of the aphasic records departed considerably from normal usage (as defined by analysis of twelve control records), with similarity among some patients in the pattern of divergence. The measures used appear to be of particular value in revealing (i) semantic difficulties in word selection and (ii) difficulties in the sequencing of speech that occur along with syntactic losses.

This paper describes some characteristics of the speech of 12 adult aphasic patients, as compared with the speech of 12 control subjects. From transcripts of subjects' speech a number of measures are determined which may be useful as general descriptions of speech and which may be sensitive to various sorts of language disorder. This information permits some evaluation of the divergence of a particular aphasic record from the control data, as well as its divergence from each of the other aphasic records.

To obtain running samples of free speech, 20 cards from the Thematic Apperception Test were administered to each subject; all responses were tape recorded and subsequently transcribed. The transcripts were processed so as to provide for each subject a list of minimal free forms (these will be called words), their sequence, and frequency of occurrence. For some purposes the transcript items were classified (with difficulty in the case of ambiguous items, particularly for the aphasic patients) in terms of the following form classes: adverb, adjective, verb, noun, pronoun, other (syntactic words: articles, prepositions, and conjunctions), neologism, unusual use (unclassifiable), period, pause, and vocal gesture (or, more simply, gesture). As the occasion warranted these form classes were collapsed into a smaller set of adverb, adjective, verb, noun, pronoun, other; or even into lexical class (adverb, adjective, verb, noun) and function class (pronoun, other).

* *This investigation was supported by PHS research grants M-1849 and M-1876 from the National Institute of Mental Health, United States Public Health Service. The authors are grateful to Dr. R. Darrell Bock, for general assistance in this project, and to Mr. Raymond A. Wiesen, who collaborated in the plans for data analysis and developed the computer programme for the analysis. At the University of Chicago, students who aided in data collection and linguistic classification include Robert C. Hunt, Janice Lynn, Norman Markel, and Doris Van Pelt.*

Essentially, three aspects of speech are examined: a) the relative frequency of use of each of the form classes, b) the concentration or diversity in vocabulary as measured by the type-token ratio or by Yule's *K* coefficient (Yule, 1944), and c) sequential dependencies in form-class usage, transitional digram frequencies as indicated by the relative frequency with which items in any one form class are followed (or preceded) by items from each form class.

Utility of the measures

A speaker is regarded as being equipped with a repertory of words, together with a set of rules for the formation of strings of words; on any particular occasion his task may be thought of as the selection of items from his word pool and the arrangement of these consistent with formation rules in terms of some plan (*cf.*, Miller, Galanter, and Pribram, 1960).

Consider first some possible consequences of losses in a person's word repertory. If losses were concentrated among the less frequent, more specific lexical items, one might expect a redistribution of usage over form classes; e.g., such losses would particularly affect availability of nouns, and we might anticipate greater use of the more frequent and general substitutes for nouns, i.e., pronouns. Such losses would also be likely to result in greater concentration or lesser diversity in use of lexical items. Those few items which remain available would be expected to be used more often. Almost necessarily, as a consequence of the above, there are likely to be changes in the contingencies among items, i.e., in the immediate environment of members of any given form class. All of these predicted effects would be reflected by changes in the three measures outlined above.

For the free-speech situation in which the data were gathered, disorders of word selection may be regarded as having much the same consequences for speech as word losses from the repertory, with effects of the sort indicated above. In other circumstances, say a structured test situation, quite different performances might result; if the word-selection process were to be disturbed, the speaker should have difficulty even if an appropriate word repertory were provided experimentally.

Losses of syntactic items, characteristically of intralinguistic function and of high frequency of occurrence in normal speech, would be reflected in other ways. But again, we would anticipate changes in form-class usage, in measures of concentration of vocabulary for function items, and, of course, changes in the sequencing or arrangement of items from the different form classes.

Disruption of the rules of arrangement for strings of words will surely be reflected in changes in word sequencing at the digram level. Just as losses with regard to the rules of arrangement for strings of words will lead to changes in sequencing, so changes in sequencing may well lead to losses or changes in the occurrence of words from given form classes, e.g., with loss or change in the formation rules that make for complex grammatical constructions there is likely to appear a concurrent decrease in

TABLE 1

CASE	AGE	SEX	EDUCATION*	OCCUPATION	DIAGNOSIS	PARALYSIS
102	40	M	I	Tool & die designer	Cerebral Aneurism	Right
112	40	M	I	Carpenter	Post-brain tumor extirpation, left tempo-parietal	Right
105	49	M	II	Clerk	Post-encephalitis, post exploratory brain surgery, left temporal	None
107	35	M	II	Semi-skilled steno-clerk	CVA	Right
109	40	M	II	Tool & die maker	Post-brain tumor, left temporal	Right
110	58	F	II	Housewife	CVA	Right
106	70	M	II	Mechanic	CVA	Right
111	47	M	III	Executive	CVA	Right
104	54	M	III	Lawyer	CVA	Right
108	35	M	III	Dentist	Trauma, accident	Right
101	65	M	III	Lawyer	CVA	None
103	67	M	III	Agronomist	CVA	None

* I: *Less than 12th grade*
II: *Completed high school*
III: *Completed college*

Aphasic patients: personal data.

frequency of the small syntactic function words which provide the scaffolding for such constructions (*cf.*, Wepman, Bock, Jones, and Van Pelt, 1956).

The measures to be considered here obviously are not the only interesting ones that might have been obtained. They may not be sensitive to certain important structural features in speech or the disruption of speech. Unfortunately, each cannot be regarded as a univocal indicator of some one underlying aspect of speech. Nevertheless it seems plausible, in terms of the considerations noted above, that some very general, important features of language or damage to language are reflected by the measures considered.

METHOD

Subjects

The twelve aphasic patients considered in this study were referred to the Speech Clinic of the University of Chicago, and there each was diagnosed as suffering from some aphasia. Background information on these patients is summarized in Table 1. The twelve control subjects were all white, native-born with English as native tongue, at least of average intelligence, without history or indication of psychotic disturbance

TABLE 2

CASE	AGE	SEX	EDUCATION*	OCCUPATION
009	36	M	I	Bus driver, Army
012	44	F	I	Dishwasher and housewife
008	91	M	I	Retired farmer, Salvation Army
011	69	F	I	Housewife
007	42	M	II	Foreman on assembly line
001	40	F	II	Housewife, saleswoman
006	73	M	II	Carpenter
010	62	F	II	Missionary, Salvation Army
005	47	M	III	Dentist
003	50	F	III	Librarian, retired teacher
004	67	M	III	Lexicographer, professor
002	66	F	III	Retired teacher

* I: *Less than 12th grade*
II: *Completed high school*
III: *Completed college*

Control subjects: personal data.

or organic brain damage. Table 2 provides a summary of information on the control subjects.

Definition of Measures and Results

Length of Transcript. The data of Table 3 allow some gross comparisons among subjects with regard to fluency or total amount of speech in response to 20 TAT pictures.

Form-Class Usage. The relative frequency of use of the various form-classes is shown in Table 3[1] where divergence of aphasic scores from the control range is also indicated.[2]

[1] *A reduced form-class usage table was also obtained where relative frequencies were computed with Period, Pause, and Vocal Gesture classes omitted. The results generally were much the same. In all cases, words were classified, except, of course, for neologisms and "unusuals" on the basis of their classification in the American College Dictionary. If, as is so common in English, more than one classification was possible, the context was allowed to decide between them.*

[2] *Generally, in evaluating data from an aphasic patient it will be indicated whether his score on the measure under consideration falls within, below, or above the range of scores of the control group (for this purpose, the control range will be defined as the range of the middle 10 of the 12 observations). In comparisons between aphasic patients, similarities in deviation from the control data will be of particular interest.*

Table 3

S	N	OTHER	NEOL.	PERIOD	PAUSE	GESTURE	ADV.	ADJ.	NOUN	VERB	PRO.	UNUSUAL
101	2664	.241	.014+	.049+	.011	.029	.092	.101+	.177+	.185−	.103	.012+
102	1025−	.061−	.013+	.037	.192+	.213+	.049−	.035−	.115	.148−	.122	.015+
103	1565−	.208	.004	.031	.109+	.073+	.089	.084	.123	.181−	.088−	.015+
104	1849−	.166−	.006+	.008−	.144+	.072	.120+	.073	.089−	.163−	.121	.038+
105	1836−	.204	.000	.060+	.044	.016	.086	.088	.146	.214	.138	.004
106	5692+	.164−	.001	.040	.100+	.028	.102	.056−	.101−	.213	.161	.039+
107	5558+	.146−	.020+	.029	.048	.101+	.143+	.043−	.064−	.193−	.181+	.035+
108	4807+	.160−	.007+	.066+	.071	.087+	.119+	.070	.103−	.182−	.124	.017+
109	3396	.163−	.001	.029	.045	.077+	.102	.074	.096−	.225	.171+	.016+
110	5524+	.156−	.010+	.035	.068	.044	.124+	.065−	.090−	.193−	.190+	.026+
111	774−	.165−	.001	.048+	.217+	.150+	.044−	.021−	.130	.125−	.061−	.037+
112	4182	.208	.022+	.034	.016	.017	.092	.091	.150	.219	.120	.033+
CONTROL SUBJECTS:												
Mean	3423	.216	.002	.031	.040	.035	.094	.087	.136	.218	.137	.003
Range:												
Middle 10	2346−	.193−	.000−	.017−	.008−	.010−	.077−	.070−	.105−	.195−	.103−	.001−
	4630	.258	.004	.040	.077	.072	.106	.097	.172	.243	.164	.005

\+ *above control range*
− *below control range*

Relative frequency of use of various form classes.

Table 4

S	K FOR LEXICAL WORDS	K FOR FUNCTION WORDS
101	132.9	656.9
102	935.1+	2797.9+
103	169.6+	690.8
104	223.2+	690.8
105	151.4	563.9
106	256.8+	432.6−
107	289.4+	621.0
108	95.5	469.7−
109	163.2+	641.4
110	198.2+	572.3
111	412.4+	1119.3+
112	87.2	510.9
CONTROL SUBJECTS:		
Mean	117.9	555.0
Range: Middle 10	74−157	475−714

\+ *above control range*
− *below control range*

Yule's coefficient K determined separately for lexical and function words.

Concentration or Diversity in Vocabulary. The type-token ratio, which may be considered a measure of diversity in vocabulary, was computed separately for lexical and function items, as well as for the total test. However, these values are extremely difficult to interpret because of their sensitivity to sample size ; generally speaking the smaller the number of words spoken the larger the type-token ratio. Because of this difficulty, no use is made of the measure here.

A particular virtue of Yule's K coefficient (Yule, 1944) as a measure of concentration in vocabulary use is that it is independent of sample size, and one may therefore legitimately compare K values from texts of different lengths. Yule's K coefficient is defined as:

$$K = (10^4)\left(\frac{S_2 - S_1}{S_1^2}\right)$$

with $S_1 = \sum_i (f_{x_i} x_i)$ and $S_2 = \sum_i (f_{x_i} x_i^2)$, where f_{x_i} is the number of words appearing x_i times. If n is the number of different words in a transcript and N is the total number of occurrences, i.e., the length of the transcript, then, for each of the n distinct words, the quantity x_i/N gives the relative frequency of use of that word. The quantity K is linearly related to the squared coefficient of variation of these n relative frequencies x_i/N. Table 4 provides the relevant information separately for lexical and function items. In Table 5 further information is to be found on diversification in use of function items.

Sequential Dependencies in Form-Class Usage. In considering first-order sequential dependencies among form classes, two distinct sets of relative frequencies may be reported. Given the occurrence of an item in a *reference form class,* say form class A, (i) we may determine the relative frequency with which such an item is *followed* by items of every form class or (ii) we may determine the relative frequency with which such an item is *preceded* by items of every form class. Since items from the various form classes do not occur equally often, (i) the relative frequency with which items from form class B follow items from form class A is not generally equal to (ii) the relative frequency with which items from form class A precede items from form class B. In the one case total number of occurrences of A form the reference basis of the relative frequencies, while in the second case B is the reference form class. In Table 6 appears the control range of relative frequencies with which items from each (reference) form class are *followed* by items from each form class. Table 7 shows the control range of relative frequencies with which items from each (reference) form class are *preceded* by items from each form class. The starred cells indicate the three form classes which most commonly follow (or precede) the occurrence of a member of a given class. For aphasic transcripts, then, one may examine departures from normal sequencing for all categories, or just for the categories most likely in the immediate environment of a member of a particular form class.

It should be noted that subjects with *similar* base frequencies of occurrence of different form classes (or with similar departures from normal in these frequencies) *can have* different patterns of digram frequencies. However, subjects with differing base

TABLE 5

7	(1)	(2)	(3)	(4)
101	25	35	29	64
102	8	11−	6−	17−
103	24	34	18−	52−
104	22	24−	23	57
105	25	31−	21−	52−
106	25	32	31	63
107	22	30−	33+	63
108	25	35	31	66
109	23	27−	27	54−
110	24	29−	30	59
111	10	17−	7−	24−
112	25	41	32+	73
CONTROL SUBJECTS:				
Mean		39.2	27.5	66.7
Range: Middle 10		32−48	23−31	57−76

(1) No. of different syntactic items (other) among 25 items used by 10 or more of control subjects.
(2) No. of different syntactic items (other).
(3) No. of different pronouns used.
(4) Total No. of different function words used (pronouns and other).

\+ *above control range*
− *below control range*

Diversity in use of syntactic items.

TABLE 6

	OTHER	PERIOD	PAUSE/ GESTURE	ADJECT./ ADVERB	NOUN	VERB	PRONOUN
OTHER	195-278*	000-007	021-122	113-228	212-329*	022-065	112-247*
PERIOD	071-304	000-000	123-613	052-129	009-032	009-090	123-358
PAUSE/ GESTURE	143-275	001-024	036-228	109-204	022-110	071-145	162-353
ADJECT./ ADVERB	125-186	014-056	014-079	172-253*	255-302*	127-184*	064-142
NOUN	256-418*	024-136	021-238*	082-150	066-156	127-174*	074-128
VERB	221-290*	016-042	019-058	224-292*	022-043	242-295*	057-135
PRONOUN	058-107*	021-047	017-059	045-082	001-011	614-712*	058-171*

* *The three form classes which most commonly follow occurrences of each (row) class.*
(Decimal points are omitted from all entries.)

Relative frequency with which item from given (row) form class is followed by item from each (column) form class—range of middle 10 control subjects.

TABLE 7

	OTHER	PERIOD	PAUSE/ GESTURE	ADJECT./ ADVERB	NOUN	VERB	PRONOUN
OTHER	195-278*	000-036	103-210	205-292*	384-492*	024-055	223-319*
PERIOD	002-058	000-000	063-337	008-024	002-006	002-009	021-152
PAUSE/ GESTURE	025-139	000-057	100-281	029-092	011-066	012-075	086-176*
ADJECT./ ADVERB	119-154	112-239	080-177	184-272*	281-422*	112-150*	080-174
NOUN	180-279*	195-536	088-223	046-101	066-156*	063-149	051-149
VERB	210-305*	122-230	083-147	278-342*	033-069	242-295*	099-210*
PRONOUN	037-068	084-415	025-082	025-067	000-009	367-451*	058-171

* *The three form classes which most commonly precede occurrences of each (column) class.* (Decimal points are omitted from all entries.)

Relative frequency with which item from given (column) form class is preceded by item from each (row) form class—range of middle 10 control subjects.

frequencies *are expected* to display differences in sequential or digram frequencies. It is obvious that the above consideration will make for difficulty in interpreting differences in digram frequencies when there are also differences in base frequencies of form-class usage. Another consideration that tends to make for difficulty in interpretation of these results also should be mentioned. Given that the relative frequencies (or probabilities) for an aphasic patient fall outside the normal range for a number of cells, it is difficult to know how to summarize these differences; the same problem arises where attempts are made to assess the similarity between patients in departure from the control data. The problem is aggravated since departures are not independent, within a row in Table 6 and within a column in Table 7. The substantive interpretation in psychological terms of departures from the normal sequencing of language would appear to be a formidable task.

Examples of the data on departures from normal sequencing are to be found in Tables 8a and 8b.[3]

Table 9 provides information on the extent of the departures from normal usage in terms of the number of divergent cells for each patient for all transitions, and for the three most common transitions for each form class. (Here the Period and Pause/Gesture categories are not considered as reference categories but only as environmental categories for the other form classes.)

Further information as to the locus of deviations from normal sequencing may be found in Table 10, where appear the categories for which transition frequencies show the largest number of departures from control usage.

[3] *The complete data on departures from normal sequencing, for all aphasic patients, may be obtained from the authors.*

TABLE 8a

	OTHER	PERIOD	PAUSE/ GESTURE	ADJECT./ ADVERB	NOUN	VERB	PRONOUN
OTHER	0	0	++	0	0	– –	–
PERIOD	– –	0	0	++	0	–	0
PAUSE/ GESTURE	–	0	++	0	+	+	– –
ADJECT./ ADVERB	0	0	++	0	–	– –	0
NOUN	0	++	0	0	0	0	–
VERB	0	0	++	0	–	– –	–
PRONOUN	– –	0	0	–	–	++	– –

++ = *above control range*
+ = *at upper border control range*
0 = *within control range*
– = *at lower border control range*
– – = *below control range*

Record No. 103 : Departures from normal sequencing of speech (for relative frequencies of different form classes *following* appearance of item from given form class, read across rows).

TABLE 8b

	OTHER	PERIOD	PAUSE/ GESTURE	ADJECT./ ADVERB	NOUN	VERB	PRONOUN
OTHER	0	0	0	0	0	–	0
PERIOD	0	0	0	++	0	–	0
PAUSE/ GESTURE	0	0	0	++	++	++	++
ADJECT./ ADVERB	0	– –	++	0	0	–	++
NOUN	0	++	–	0	–	0	0
VERB	0	0	++	– –	– –	– –	0
PRONOUN	– –	– –	0	–	0	0.	–

++ = *above control range*
+ = *at upper border control range*
0 = *within control range*
– = *at lower border control range*
– – = *below control range*

Record No. 103 : Departures from normal sequencing of speech (for relative frequencies of different form classes, *preceding* appearance of item from given form class, read down columns).

TABLE 9

S	A-FOLLOWING GIVEN FORM CLASS (1) *Max.* = 35	(2) *Max.* = 15	B-PRECEDING GIVEN FORM CLASS (1) *Max.* = 35	(2) *Max.* = 15	A+B (1) *Max.* = 70	A+B (2) *Max.* = 30
101	4	1	5	1	9	2
102	21	12	21	8	42	20
103	10	5	11	2	21	7
104	12	7	13	9	25	16
105	5	3	2	2	7	5
106	10	6	10	7	20	13
107	13	6	16	9	29	15
108	7	3	6	2	13	5
109	6	3	4	3	10	6
110	10	5	11	7	21	12
111	28	12	22	10	50	22
112	1	1	2	0	3	1

(1) — *all divergences.*
(2) — *divergences among the three most likely transitions for each form class.*

(Period, pause/gesture categories omitted as reference categories, kept as environmental categories).
Transition relative frequencies : total number of divergent cells.

Ratings by Linguist. Having read the transcripts of the aphasic patients, a linguist judged them with regard to their similarity to each other and the nature of the departures from normal, the extent to which divergencies were of a semantic kind (difficulties in word selection), or a syntactic kind (disruptions of the grammatical matrix of language), or of a pragmatic sort (where disruption of the integrative processes in language formulation leads to speech which conveys no meaning even though syntactic structure is largely retained). The linguist's judgments of the aphasic records are schematically shown in the spatial model of Figure 1.[4] For discussion of the semantic, syntactic, and pragmatic processes, see Wepman, Jones, Bock, and Van Pelt (1960).

DISCUSSION OF RESULTS

Departure of Aphasic Records from Normal.

It is clear that in various ways the speech of nine of the twelve aphasic patients is noticeably different from normal (all except Subjects Nos. 101, 105, and 112). For each form class at least half of these nine transcripts diverge from normal in terms of frequency of usage (Table 3). Eight show greater than normal concentration in

[4] *We are grateful to Mr. Morris F. Goodman, University of North Carolina, for judging these transcripts.*

TABLE 10

S	FOLLOWING GIVEN FORM CLASS	PRECEDING GIVEN FORM CLASS	BOTH
101	None	Adj/Adv 3	None
102	Other 5, Adj/Adv 5, Noun 4, Verb 4, Pro 3	Other 4, Adj/Adv 6, Noun 4, Verb 5	Other, Adj/Adv, Noun, Verb
103	Pro 3	Pro 3, Adj/Adv 3	Pro
104	Adj/Adv 4, Verb 4	Other 5, Adj/Adv 4	Adj/Adv
105	None	None	None
106	Other 3, Adj/Adv 4, Verb 3	Adj/Adv 4, Verb 3	Adj/Adv, Verb*
107	Adj/Adv 4, Verb 4	Adj/Adv 4, Noun 4, Verb 4	Adj/Adv, Verb*
108	Noun 3	Adj/Adv 3	None
109	Other 3	None	None
110	Noun 3, Verb 4	Verb 5	Verb
111	Other 6, Adj/Adv 6, Noun 3, Verb 6, Pro 7	Other 5, Adj/Adv 6, Noun 3, Verb 5, Pro 4	Other, Adj/Adv, Noun, Verb, Pro
112	None	None	None

(Period, pause/gesture omitted as reference categories).

* The fact that two records show that particular categories carry a heavy share of departure from normal does not indicate that these departures are in the same direction.

Transition relative frequencies : form classes with three or more transition relative frequencies outside normal range.

use of lexical items (Table 4). Four depart from the control range in terms of concentration of use of function items (Table 4). Considering the matrices of transition relative frequencies for these nine subjects, the number of cells which depart from the normal varies from 10 to 50, where total number of cells is 70 (Table 9). To cite one further instance (from Table 10), there is no form class whose immediate environment is not considerably changed in one or another of these records.

It might be noted from Table 3 that there is indication of a substitutive relation between Pause and Gesture—with the exception of Subjects 102 and 111, patients over-using one will tend not to over-use the other and vice versa—and that for most of the aphasic patients, items from these categories occur frequently (more than for the normal subjects). They occur particularly in encoding immediately after frequent, low-informational syntactic items and preceding difficult semantic choices of high informational value. There are some data (Osgood, 1957) which indicate that for normal individuals both filled and unfilled pauses " tend to occur just *after* simple form class words . . . and just *before* lexical items " ; such an effect seems to be exaggerated for aphasics.

At this point a caveat with regard to the control data is both appropriate and necessary. Inspection of Tables 1 and 2 shows that on some selection variables the

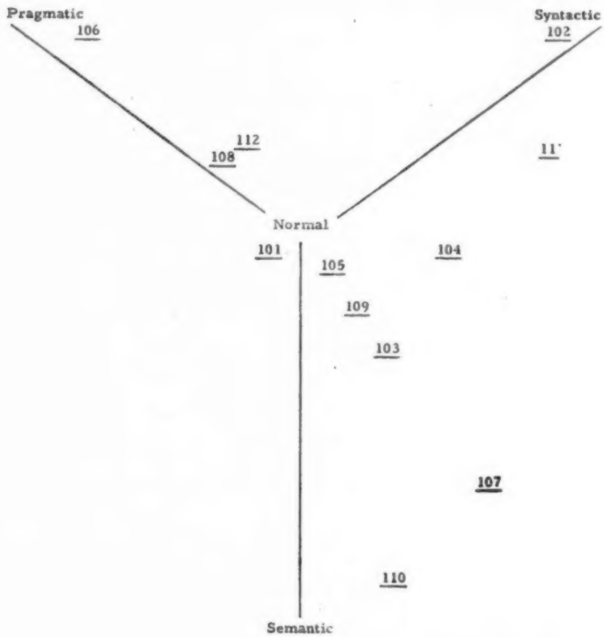


Fig. 1. Characterization of aphasic transcripts by a linguist—classification with reference to semantic, syntactic, and pragmatic difficulties.

controls differ from the asphasics (e.g., half the control subjects are women and only one of the aphasic patients is a woman) and that the control subjects were systematically selected in terms of educational level, age, and sex. This suggests that for some purposes contrasts between aphasics and only certain control subjects might have been appropriate (though this would have meant a base comparison level defined in terms of very few control subjects indeed). Since no systematic differences are apparent, related to sex or age of control subjects, it may safely be judged that contrasts between aphasic patients and the control group are of such magnitude as to transcend the relatively more subtle differences among normals.[5]

Similarities among Aphasic Records in Departure from Normal.

We now turn to a somewhat different question, namely the extent to which we can identify sub-groups of aphasic patients in terms of departure of speech character-

[5] *Analysis of the 12 normal transcripts does suggest slight differences related to educational background of subjects. Those subjects who failed to complete high school tend to differ from better educated subjects in their increased use of pauses and pronouns, their decreased use of " other," syntactic, words.*

istics from those of the normal sample with regard to stereotypy in vocabulary, form-class usage, and sequential dependencies in form-class usage. A number of records, those of patients Nos. 101, 105, and 112, show little departure from normal usage, a number of records, those of patients Nos. 107, 109, and 110, give evidence particularly of semantic difficulties in word selection, while records Nos. 102 and 111 evidence syntactic losses with difficulties especially in the arrangement and sequencing of speech. The remaining records show a variety of difficulties and do not closely resemble each other or the records mentioned above.

Records Nos. 101, 105, and 112 do not differ very much from normal, and are not very different from each other. With regard to these patients the findings may be interpreted as indicating either that, in fact, the speech of these persons has only been little impaired, or that our measures are insensitive to the kind of damage involved. A reading of these transcripts (and the judgments of a linguist based on such a reading) suggests that the former is the more plausible alternative, at least for records Nos. 101 and 105, and therefore these records will not be considered further.

In a number of ways records Nos. 102 and 111 resemble each other. Both are very short and exhibit over-use of pauses and gestures. After omitting occurrences of periods, pauses, and gestures, the adjusted relative frequencies of use of Adjective and Noun categories are exceptionally high for both records. In each case there is over-concentration in use of both lexical and function forms (these are the only records with over-concentration in both lexical and function items) and very few different syntactic items are available[6]; both records show many divergences from normal with respect to digram frequencies, the environment of almost every form class being severely disturbed. Yet in some ways the records differ. In terms of adjusted relative frequencies (omitting Period, Pause and Gesture categories), No. 102 under-uses syntactic items and over-uses pronouns while No. 111 tends to under-use pronouns and to over-use syntactic items; and as noted before, there are considerable differences in regard to between-form-class transitional frequencies. In terms of the severe disruptions of the transition frequencies, of the paucity in number of different function items, the relative over-use of nouns, and the meagre length of the transcripts, that which is common to the disorders of these two might be specified as a contiguity or arrangement defect, primarily affecting syntactic sequencing processes (see Jakobson and Halle, 1956; also Wepman, Van Pelt, Jones, and Bock, 1956).

Records Nos. 107 and 110 (and to a lesser degree No. 109) are similar in their departure from normal: these records are rather long; the patients under-use the syntactic words (falling in the Other category), under-use adjectives and nouns, and over-use pronouns and adverbs; they show greater than normal concentration for lexical but not for function items; they show considerable departure from normal with regard to transition frequencies and display disruptions in the environment of a number of form classes, in particular the environment for verbs. Still, there remain some differences between the two with regard to transition frequencies and the form

[6] Of the 25 most common Other items used by 10 or more of the control subjects, 102 uses only 8, and 111 only 10, while no other aphasic patient uses less than 22 of these items (Table 6).

classes whose environment is most disturbed, e.g., the environment for modifiers is much more disturbed in record No. 107 than in No. 110. On a variety of grounds one might regard these patients as suffering primarily from semantic, selection difficulties: diversification in use of lexical items is less than normal, with under-use of nouns and adjectives, the more specific, less frequent, high-informational, lexical items, and with substitutive over-use of pronouns, which, of course are words of high expected frequency in normal speech. With respect to these characteristics, the protocols resemble the illustrative case of a semantic aphasic presented previously (Wepman, *et al.*, 1956). However, this is clearly only a partial account, since these patients also show losses of syntactic items (under-use of the Other category), and at least some of the considerable departure from normal with regard to transition probabilities may be attributed to such syntactic losses and difficulties in the sequencing or arrangement of words.

A somewhat different interpretation is also possible for patients 107 and 110. Syntactic items must be learned individually in an intra-linguistic fashion and are not generally mutually substitutive. If items which have few or no synonyms or adequate alternatives are regarded as " specific " items, then under-use of the Noun category and of the Other category may both be viewed as indicating losses in " specific " items, and difficulties in arrangement and changes in transition probabilities may be seen as consequences of a scarcity of items that provide the frame for the longer more complicated grammatical constructions.

The remaining aphasic transcripts seem to closely resemble neither each other nor the ones already discussed.

Record No. 103 is short, involves much use of pauses and vocal gestures, and shows no departure from normal with regard to relative frequency of form-class usage, after omission of Period, Pause and Gesture categories. There is some over-concentration in use of lexical but not of function items, and evidence of restriction in the number of different pronouns available ; a considerable number of transition relative frequencies depart from normal, and the environment of the Pronoun class is particularly disturbed. A reading of this transcript reveals that it makes sense, that this patient does communicate fairly adequately but that there are considerable word-finding difficulties. More often than not the word eventually is found but only after considerable hesitation. This behaviour is consistent with the patient's over-use of the Pause and Vocal Gesture category, which is particularly evident after syntactic and before lexical items, i.e., at points of much uncertainty where difficult semantic choices or selections must be made. However, there is also considerable hesitation before the occurrence of pronouns. Perhaps this patient when faced with a difficult semantic noun choice solves the problem, after hesitation, by use of a more common pronoun. There is some over-concentration in use of lexical items, and this too might be regarded as evidence of difficulty in word selection.

Record No. 104 is short with considerable evidence of hesitation. There is some over-use of the Adverb class and some under-use of nouns and syntactic items ; there is over-concentration in the use of lexical but not of function items ; a large number of transition probabilities depart from normal and the environment of a number of form

classes is disturbed, the environment for modifiers being most severely affected. Inspection of the transcript shows considerable use of unfinished constructions, which might be correct if completed, and between-phrase discontinuity of two kinds: discontinuity because of a lack of proper syntactic connective items, and discontinuity because of the insertion of phrases having little referential content which are interspersed between the phrases carrying the communication, as though the patient were suffering from difficulties in word and phrase finding. It appears that a number of difficulties are involved—there are contiguity difficulties in word sequencing and there are difficulties in selection of the more specific informational items. To some extent the semantic selection difficulties are reflected in a tendency toward under-use of nouns and over-concentration in use of lexical items. To some extent the sequencing difficulties are reflected in the departure of transition frequencies from normal and in frequent hesitation; however, these measures can hardly be said to reflect just these difficulties, and it is not at all clear how any measure or combination of these measures reveals the difficulties in phrase finding.

The transcript of patient No. 106 is long. There is some under-use of adjectives and syntactic items together with greater than normal diversification in the function items used. Yet there is greater than normal concentration in use of lexical items. Many of the transition probabilities depart from normal and the environment of a number of form classes, particularly that for modifiers and verbs, is considerably disturbed (being altered in ways different from the disturbances in the case of No. 107). Even on a very careful reading it is difficult to make much sense of this transcript. The ratings of the linguist indicate speech grossly disturbed, with losses primarily of a pragmatic sort.

The speech of No. 106 differs from normal and the other records in a number of respects, viz., the under-use of syntactic items in conjunction with over-diversification of those function items used and over-concentration in lexical items. From a subjective assessment of the transcript, such disturbances are not totally unanticipated. For, in speech of this sort, there may be much discontinuity at "phrase" boundaries, and if phrases are short there should be considerable departure from normal with regard to transition probabilities; also, if there is discontinuity between phrases the syntactic items connecting these may be used relatively promiscuously, unselectively, and therefore more equally, resulting in a relative over-diversification in use of function items.

The transcript of patient No. 108 is somewhat lengthy, with considerable use of vocal gestures. There is some tendency to over-use adverbs and to under-use syntactic items; there is somewhat greater than normal diversity in use of function items, and while some of the transition frequencies depart from normal, these departures are less extensive than those for most of the other patients, and there are relatively few severe disruptions of the environment for any form class. Inspection of the transcript reveals difficulty in word finding and, perhaps more important, considerable discontinuity at the phrase level with much intrusion of material that stands in no obvious sensible relation to the text that comes before it or after it, losses of a sort that we have called

pragmatic. It is apparent that the measures obtained are not very sensitive to such losses. This point is illustrated again by the fact that transcript No. 112, which in terms of these measures falls in the cluster very close to normal, also shows evidence of pragmatic difficulties, though perhaps to a lesser extent. (As may be recalled, the linguist judged records Nos. 108 and 112 to be close together, noting difficulties of a pragmatic kind.)

It is of interest to note some points of agreement and disagreement between Figure 1, representing a linguist's ratings of the various aphasic protocols, and the findings presented above. As to points of agreement: records Nos. 101 and 105 are seen as falling fairly close together and closest to normal ; No. 112 is not too far removed from these, and not too far from the origin ; No. 102 and No. 111 are seen as falling fairly close together and quite removed from normal with divergences particularly of a syntactic sort ; Nos. 107 and 110 fall closest together, considerably removed from normal with semantic losses (see the data on under-use of nouns and over-use of pronouns by these patients) ; and No. 106 is seen as very far from the origin, and far from almost all the other records. As to points of partial disagreement: while Nos. 104 and 108 differ considerably in their placement on Figure 1, these patients are rather alike in form-class usage although they do differ considerably in transition probabilities ; No. 103 is rather different from Nos. 101, 105, and 112 in Figure 1, which is not consistent with the similarity in form-class usage after omission of Period, Pause and Gesture categories although it is compatible with the transition-frequency data which show this record rather distinct from Nos. 107 and 110.

The commentary on the various aphasic subjects presented above must make it obvious that, characteristically, a patient's speech will be found to suffer not from just one sort of loss or difficulty, but rather from a variety of difficulties, some of the functions important in expression of language being more disrupted than others. While it is clear that the measures obtained do not tap all the determiners of adequate speech, and that no measure stands in unique one-to-one relation to some particular underlying function, the measures do seem to reflect difficulties in the word-selection and sequencing processes (sometimes called similarity and contiguity disorders), processes which constitute some of the necessary conditions for adequate speech.

Suggestions for further research

None of the measures used in this study seems particularly valuable as an index of pragmatic difficulties, nor does the use of information analysis, or analysis of frequency of word usage seem very promising, since in an earlier paper (Van Pelt, *et al.*, 1958) these measures did not sharply differentiate an extreme pragmatic case from a normal control. It is possible that the unit of analysis employed in this paper is inappropriate for detection of pragmatic disturbances. Analysis of speech at the phrase level may become necessary to examine the connections between the semantic content of adjacent phrases and the appropriateness of the semantic content of any one phrase to the

ostensible subject matter of the discourse, or to the circumstances provoking speech. Use of the " cloze " technique (Taylor, 1953) for reconstructing mutilated utterances might yield something of interest ; for example, deletion of every third or fourth item of high semantic value (noun or verb, for example) should make the transcript of an aphasic patient who suffers from pragmatic difficulties particularly hard to reconstruct, for in this case there will be little redundancy to help the judge in attaining even the general semantic region of the missing items. A low reconstruction score resulting from this procedure, together with relatively normal scores on measures of the sort discussed in this paper might be diagnostic of pragmatic difficulties. Another possible approach might involve the use of more structured test situations requiring of the patient verbal and non-verbal plans of different complexity and scope. Evaluation of patients' performance on such tasks might serve to distinguish various sorts of speech disturbances. In this regard, and more generally also, it would be essential to obtain both free-speech data and data on performance in such a series of tasks from the same set of patients.

An experimental test approach might also be of use in clarifying some issues in the interpretation of difficulties in word arrangement (contiguity disorder). One possible explanation of this disorder is that with difficulties in sequencing there will be a loss in use of most complex grammatical constructions, with consequent under-use of the syntactic items that provide the frame for such constructions. In this paper the possibility has been raised that the primary loss may be one of syntactic, intralinguistically acquired, items ; once these are lost the use of any elaborate grammatical construction necessarily becomes impossible for the patient. An experimental situation might be designed to provide syntactic items to patients whose free speech shows syntactic loss, requiring them to use these items in building sentences. The subjects' performance might provide pertinent information which would permit some choice between the interpretations offered above.

Conclusions

A number of measures generally descriptive of language which at the same time may be sensitive to various sorts of damage to speech were used to assess some features of the free speech of twelve aphasic patients, the transcripts of these patients, elicited by TAT cards, being compared with those of twelve control subjects. In the main, three aspects of speech were examined: the distribution of words according to their grammatical function (form-class usage), sequential dependencies in form-class usage, and stereotypy in vocabulary.

In terms of the measures obtained, eight or nine of the aphasic records departed considerably from normal usage. Inspection revealed similarity among some patients in their pattern of divergence from normal. Some records showed particularly severe difficulty in word selection, other records revealed impairment especially in the arrangement and sequencing of speech with considerable syntactic losses. However, it was clear that, characteristically, a record showed not just one sort of loss, but rather a number of different kinds of departures from normal.

The measures used do appear to be of value in revealing semantic difficulties in word selection, and difficulties in the sequencing of speech that occur along with syntactic losses. The measures do not seem of much value in detecting what have been called pragmatic losses, in which case speech is almost incomprehensible because of discontinuity between successive phrases which have little semantic relation to each other. To detect such losses, analysis at a phrase level rather than the word level used in this study may be necessary.

Some suggestions for further study were made and, in particular, stress was placed on the advisability of obtaining both free-speech data and experimental test data for each aphasic patient.

References

JAKOBSON, R., and HALLE, M. (1956). Fundamentals of Language (The Hague).

MILLER, G. A., GALANTER, E., and PRIBRAM, K. (1960). Plans and the Structure of Behavior (New York).

OSGOOD, C. E. (1957). Motivational dynamics of language behavior. In Jones, M. R. (Ed.) Nebraska Symposium on Motivation: 1957 (Lincoln), 348-424.

TAYLOR, W. L. (1953). "Cloze procedure": a new tool for measuring readability. *Journalism Quart.*, 30, 415.

VAN PELT, D., WEPMAN, J. M., BOCK, R. D., and JONES, L. V. (1958). Differential disruption of symbolic processes in aphasic language disorders. Res. Rep., Psychometric Laboratory, Univ. North Carolina, October 31, 1958.

WEPMAN, J. M., BOCK, R. D., JONES, L. V., and VAN PELT, D. (1956). Psycholinguistic study of aphasia: a revision of the concept of anomia. *J. Speech and Hearing Disorders,* 21, 468.

WEPMAN, J. M., JONES, L. V., BOCK, R. D., and VAN PELT, D. (1960). Studies in aphasia: background and theoretical formulations. *J. Speech and Hearing Disorders,* 25, 323.

YULE, G. U. (1944). The Statistical Study of Literary Vocabulary (London).

LISTENER COMPREHENSION OF SPEAKERS OF THREE STATUS GROUPS*

L. S. Harms
Louisiana State University

Listeners of three statuses attempt to reconstruct spoken messages of speakers of three statuses. Speakers of high status were most comprehensible. However, listeners achieved highest relative comprehension when speaker and listener status coincided.

Introduction

Ways of talking, speech patterns, or more exactly, status dialects, are thought to develop through social group membership. Bloomfield (1933) insists that a speaker talks more like those people he talks with most frequently and less like those individuals he talks with least frequently.

Putnam and O'Hern (1955) demonstrated that untrained listeners are able to identify the social status of a speaker from voice recordings. These listener identifications correlated 0.80 with the Warner *Index of Status Characteristics* which yields scores based on a weighted assessment of the factors of sources of income, value of home, education and occupation. Ladefoged and Broadbent (1957) report that subjects of different social groups achieved different scores on a word identification test.

Spencer (1957) suggests that it is at "... those points where linguistic phenomena most closely impinge or appear to form an integral part of general human behaviour that studies of language are most needed today." The present study attempts to determine the effect of status features in speech on the comprehension of a spoken message.

Procedure

Nine speakers were employed in this study. They were male, between 30 and 50 years of age. They had lived in Ohio or a bordering state during their youth and had lived in this area most of their adult lives.

Each speaker recorded a short "advice giving" narrative. In a pre-recording interview, the speaker was asked to name one magazine, one newspaper, and one

* *Based, in part, on a Ph.D. dissertation directed by Franklin H. Knower at The Ohio State University.*

television show he thought a young boy of twelve years should know about. These titles were written on a card and returned to the speaker for his reference while making the recording. The speaker was then instructed to "pretend" that he was advising a twelve year old boy to read the suggested newspaper and magazine and to watch the television show. The speaker was asked to talk "conversationally" for two minutes. With no further rehearsal or instruction, the recording was made. All speakers talked for about two minutes and produced between 150 and 300 words.

The recordings were prepared for the *Cloze Procedure* test of listener comprehension. The literature suggested (Taylor, 1953, 1956) and pilot studies supported the view that 100 word samples were sufficient for experimental purposes. All recorded speech samples were edited to between 100 and 115 words. Editing was performed by cutting out selected complete "sentences" which did not destroy continuity of the narrative. All edited samples, in terms of context, contained mention of one magazine, one newspaper and one television show.

After the editing, 10 trained judges of oral style heard both the unedited and edited recordings. They were asked to judge the degree to which the style had been preserved in the editing process. On a nine point scale, where 1 equalled "style preserved" and 9 equalled "style distorted", an average mean of 1.8 with standard deviations ranging from 0.46 to 1.17 was found. It was concluded that style represented in the edited version of the sample was adequate for the purpose of this study.

The edited speech samples were then copied from the recording and printed forms were prepared. On these forms every fifth word was replaced by a twelve space line. This line served as the write-in blank which the listener filled in or "clozed". Each of the nine forms contained twenty write-in blanks. One of the forms used in the study is shown below.

> __________think the first thing__________
> might look at is__________New Yorker. Begin
> with__________cartoons because they are__________.
> And, then, I think__________second thing to look
> __________is some straight reporting__________—
> like the accounts of__________breaks and shop-
> lifting,__________that sort of thing.__________
> finally, the fiction—though__________find that the
> least__________part of the New__________altogether.
> As for a__________—there's only one good__________
> __________and that's the New__________Times. The
> only television__________I look at practically
> __________Playhouse 90, and it__________me as
> being better than a lot of other TV programmes just
> because it's longer.

Immediately after hearing a recording, the listener was given the appropriate form to complete, or "cloze." The listener's task was to write-in the exact word

the speaker had used during the recording. Only the exact word the speaker had used was considered acceptable for scoring purposes.

Both listeners and speakers were classed into status groupings by use of the Hollingshead Two Factor Index of Status Position (1957). The two factors were education and occupation. This simplified index correlates highly with other more complex instruments employing difficult to assess variables such as value of home, and total source of income.

The listening task was completed wherever groups of five to ten listeners could be assembled. Fire houses, living rooms, and church basements are typical of the listening environments. Only responses from those listeners indicating they clearly heard the recordings were retained in this study.

Each listener heard three speakers of which one was high status (HS), one middle status (MS) and one low status (LS). Each speaker was heard by 60 listeners of which 20 were HS, 20 were MS, and 20 were LS. Thus, a total of 1200 word write-ins were obtained for each speaker. The 180 listeners were selected from the non-college population. They were mainly employed adults and housewives.

In general, HS listeners made fewest errors, LS listeners made most errors. However, it was the frequency of errors a listener of a given status group obtained after hearing speakers of different statuses that appeared to yield the greatest information. Correspondingly, rank of errors was used as the measure. For instance, if a HS listener in "clozing" the blanks for the HS speaker made three errors, for the MS speaker five errors, and for the LS speaker eight errors, the HS speaker was assigned rank 1, the MS rank 2, and the LS rank 3. No ties were permitted. In this manner, the data presented in Table 1 were obtained.

Data were punched on IBM cards and processed at The Ohio State University Research Center on an IBM 650 computer.

Results and Discussion

Two trends in the data emerge when the results are considered. First, as illustrated in Table 1, HS speakers were the more comprehensible. Second, listeners hearing a speaker of their own status comprehended this speaker more readily than did listeners

Table 1

Mean Rank of Errors

Listeners	*Speakers*		
	HS	MS	LS
HS	1.50	1.90	2.60
MS	1.65	1.55	2.80
LS	1.85	1.83	2.32

TABLE 2

MEAN RANK OF ERRORS

	Speakers HS	MS	LS
Speaker and Listener Status Coincide	1.50	1.55	2.32
Speaker and Listener Status do not Coincide	1.75	1.87	2.70

of other statuses. The first of these trends relates to the speaker and his way of talking. The second relates to the listener and his relative success in listening to speakers of different status backgrounds. These results may be interpreted as indicating that some ways of talking are more comprehensible than others, while skill in listening is related to the listener's social status.

The expected results were that when speaker and listener status coincided, highest comprehensibility would result. Considered from the listener's point of view, this did occur, as shown in Table 2. Listeners hearing a speaker of their own status achieved higher comprehensibility than when speaker-listener status did not coincide. This finding was predicted by Bloomfield's descriptive analysis and *Cloze Procedure* rationale.

SUMMARY AND CONCLUSIONS

180 listeners of varying statuses heard short recorded messages produced by nine speakers of different status backgrounds. Short speech samples were mutilated according to *Cloze Procedure* technique. From listener attempts to reconstruct the original speaker message, the following conclusions are framed.

1. Speakers of high status are, on the average, more comprehensible.
2. Listeners more successfully comprehend speakers of their own status than do listeners of other statuses.

REFERENCES

BLOOMFIELD, L. (1933) Language (New York).

HOLLINGSHEAD, A. (1957) Two Factor Index of Social Position (printed privately).

LADEFOGED, P. and BROADBENT, D. (1957) Information conveyed by vowels. *J. acoust. Soc. Am.*, **29**, 98.

PUTNAM, G., and O'HERN, E. (1955) The status significance of an isolated urban dialect. *Language* **31**, 4 (part 2, supplement).

SPENCER, J. (1957) Received pronunciation: some problems of interpretation. *Lingua* **7**, 7.

TAYLOR, WILSON (1953) "Cloze Procedure: " a new tool for measuring readability. *Journalism Quarterly*, **30**, 415.

TAYLOR, WILSON (1956) Recent developments in the use of "Cloze Procedure". *Journalism Quarterly*, **33**, 42.

CONGENITAL LANGUAGE DISABILITY AS A STUDY MODEL OF EVOLUTION IN COMMUNICATION*

Godfrey E. Arnold**
New York

The relation between various forms of congenital language disability and laterality are reviewed in detail and the problem of cerebral dominance in man and its influence on linguistic activity is considered. The relevance of results with delayed auditory feed-back to such problems is discussed and the theory is presented that language disorders are the result of deficient homeostasis.

Definition

The complex problems of familial or developmental language disability have lately attracted much interest. Many authors feel (see Arnold, 1960) that the developmental language disorders are related to congenital aphasia, a term rejected by others for semantic reasons. To avoid this criticism, Wood (1959) selected a conservative definition : *Language disorder* is the inability of a child to use symbols for communicative purposes, resulting from injury to, or lack of development of the cortex. If we follow this definition, we may say that congenital language disability retards the development of the capacity to use symbols and their syntactic combinations for communication.

There is substantial evidence for the assumption that congenital language disability represents a *familial, constitutional, and hereditary disability.* In several papers Eustis (1947) described the hereditary syndrome of congenital dyspraxia, familial sinistrality, and congenital language disability with specific dyslexia and tachyphemia (cluttered speech). Thus, congenital or specific language disability is a special form within the general group of language disorders. Though being related to the other aphasiform types of delayed language development, it differs from the pre-, intra-, or post-natally acquired aphasias in many respects. Hence, we must distinguish two main types or forms of language disorder: (1) *Specific* or congenital language disability

* *From the Diagnostic Services (G. E. Arnold, M.D., Clinical Director) of the National Hospital for Speech Disorders in New York (Lynwood Heaver, M. D., Director). This Study was aided by a Research Grant from the Wenner-Gren Foundation of New York.*

** *Director of Research, Section of Otolaryngology, New York Eye and Ear Infirmary.*

as a familial, hereditary and idiopathic syndrome ; and (2) *Symptomatic* or acquired language disorder resulting from any brain lesion before, during, or after birth, such as in infantile aphasia due to some paratypical pathogenic cause (malformation, trauma, infection, etc.).

Hereditary transmission of general language disability manifests itself by the familial occurrence of language disorders in various combinations : delayed onset of speech development *(prolonged alalia)*, articulatory disorders *(infantile dyslalia)*, delayed acquisition of grammar and syntax *(dysgrammatism)*, reading and writing disability *(dyslexia and dysgraphia)*, often associated with delayed psycho-motor maturation (*congenital dyspraxia*), retarded differentiation of lateral dominance (*familial sinistrality*), and other signs of *delayed neural maturation.* All observers agree that each of these developmental language disorders is about four times more frequent in the male sex than among females. This is another indication of their constitutional etiology. Further evidence for the genetic nature of these familial syndromes of delayed psychosomatic maturation will be discussed under the headings of these categories.

As is well-known since the first contributions to speech pathology (see Arnold, 1960), congenital or specific language disability comprises a chronological *sequence of developmental disorders* of language and speech. Thus, a physically normal three-year-old child may first be brought for evaluation of his delayed speech development (*prolonged alalia*). In view of the general immaturity of such children, we reassure the parents with appropriate counselling and advise their return after the lapse of a year. Upon the child's second visit at the age of about 4 years, we usually see the state of *infantile dyslalia* in its severe forms (vowel speech, idiolalia, verbal dyslalia). Speech therapy is then instituted, and after a further year has gone by the child may be seen in the stage of *residual dyslalia* which affects the motorically difficult sounds of /l/, /r/ and the sibilants. *Dysgrammatism* is frequently present and lasts much longer than it does in the physiological forms of infantile speech. After formal education has begun, *congenital dyslexia and dysgraphia,* or specific reading and writing disability, interferes with school progress. At the same time, the frequent tendency of these children towards rapid and hasty speech *(tachylalia)* begins to create an increasing handicap in oral expression. At the ages around the tenth year, the symptom complex of cluttering increases in severity and the children are brought back because of their poorly formulated *(paraphrasic)* and carelessly articulated *(pararthric)* speech. During puberty, the hurried confusion of cluttered speech usually gets worse and becomes further aggravated by the insecurity of adolescent behaviour. For this reason, concerned parents may again seek assistance when the fully developed condition of *tachyphemia* seems to jeopardize higher education. Without proper treatment, cluttering persists throughout life, for it will not be outgrown.

Congenital dyspraxia

This psycho-motor disability constitutes *a developmental lag* in the acquisition of motor co-ordination. It has been described under various names: developmental

apraxia, motor infantilism, habitual clumsiness, developmental awkwardness, delayed motor maturation, and the like. Orton (1937) offered the most detailed description and mentioned that Galen knew the condition, writing that some children are ambilevous, that is doubly left-handed. This immaturity in psycho-motor development and its familial occurrence in close association with congenital language disability has been analyzed by many authors. In particular, the awkward posture and gawky motor performance of the cluttering adolescent has been regularly commented on.

Clinically, the syndrome of congenital dyspraxia appears in the following *symptoms or anamnestic data*: (1) Late motor maturation with delayed ability to sit, stand, walk, hop, climb, etc.; (2) Delayed onset of speech development; (3) Unusual clumsiness of all motor performances for sports, manual activities, social dances, etc.; (4) Late establishment of cerebral dominance and kindred distortions of spatial relationships, especially of the lateral directions to right and left.

According to Karlin's theories (1958), this syndrome may be related to *delayed myelinization* of the motor neurons. When following such theories further, it does not appear unreasonable that their application to an etiological explanation of stuttering may actually be valid for the psycho-motor genesis of cluttering and its potential complication by stuttering. It follows that the concept of congenital dyspraxia may be helpful in the differentiation between the two main types of stuttering, namely the *idiopathic* form with good language ability, and the *tachyphemic* form on the basis of a primary language disability.

Dyspractic articulation, poorly co-ordinated movements of phonic respiration, and a dysrhythmic quality of all somato-motor actions reflect a basic disorder of neuro-muscular balance. Seeman (1959) interprets this imbalance as an organic dysfunction of *extrapyramidal co-ordination* which is furthermore complicated by a lack of cortical inhibition. Since the regulations of all rhythmical and periodic body functions (heart rate, respiration, locomotor patterns, sleep-wake periods, digestion, etc.) take place in the thalamic and subthalamic vital centres, it is not surprising that the motor elements of speech should be influenced by all states of extrapyramidal function. This truism is demonstrated by the well-known pathology of extrapyramidal dysarthria, such as occurs in Parkinsonism. In fact, certain extrapyramidal dysarthrias may be so similar to cluttering that careful differential diagnosis is always necessary.

The problems of lateral dominance

The influence of left-handedness on the occurrence of developmental language disorders and their end stage of cluttering is little understood although many authors stress the frequency of this association (see, e.g., Orton, 1937). Since we know that familial sinistrality and mixed dominance are undoubtedly associated with congenital word-blindness, particularly with the *strephosymbolic form of dyslexia,* it would be premature to reject a similar correlation with the cluttering language debility. More clarity may result from a carefully planned analysis of clinically determined dominance

in correctly diagnosed clutterers. However, it seems to us that disturbed dominance may be associated only with certain forms of cluttering, namely, in those cases which are combined with specific dyslexia.

The more I read about the mythically tinged problem of lateral dominance, the more I get confused by the increasing number of theories encountered. Whenever a subject is discussed by first listing a long series of theories and by then adding the author's own theory, we can be sure of the generally limited understanding of that subject. Since lateral dominance in the function of brain, hands, feet, eyes, and possibly ears, is the only purely human acquisition which has no parallel in animals, all related problems, such as language, speech, writing, reading, and preferential usage of hands, are discussed by specialists in widely varied fields of knowledge. Neurologists, psychiatrists, psychologists, physiologists, pedagogues, graphologists, anthropologists, phoneticians, linguists, phoniatrists, pediatricians, physiatrists, and even theologians analyze the obscure facts and fancies about handedness from their specific viewpoints. The resulting conclusions are obviously influenced by the additional problem of semantics.

Although Kainz's encyclopedic opus (1954 - 1956) on the psychology of language has greatly promoted the clearer understanding of many concepts pertaining to the semantic structure of communication in all its ramifications, scientific terminology of speech and language pathology is still far from being exact, standardized, and generally accepted. Consequently, one encounters a great confusion of concepts, terms, and definitions in many contributions to the problems of laterality and language. Unfortunately, the concepts of language, speech, writing and reading, and particularly the countless possibilities in the derangement of these different functions are used with insufficient or even false differentiation. For instance, the fundamentally different communication disorders of stuttering dysphemia, cluttering tachyphemia, psychogenic mutism, articulatory dyslalia, and specific dyslexia are frequently misinterpreted as one single entity of " developmental speech disorders ".

Such misunderstandings are as serious as if someone claimed that all respiratory disorders were allergic (which is true for asthma), or infectious (which is true for tuberculosis or pneumonia), or cardiac (which is true for cardiac failure), or occupational (which is true for silicosis or anthracosis), and so on. Before modern medicine had learned the *patho-anatomical classification* of diseases by etiological agents and pathological changes, the *symptomatic diagnosis* of medieval physicians and surgeons relied on intuitive speculations on the magic influences of miasmas, elements, or noxious air (hence the name " malaria " for infection with plasmodium transmitted by the anopheles mosquito).

This, then, is our present dilemma. In spite of many valuable contributions, no definite answers are available for the explanation of the *origin of manual and cerebral dominance in man,* its ultimate causes, and its influence on language in general. If we try to state the presently accepted facts, we come to the following conclusions.

Disorders of language and laterality

(a) *General concepts of laterality* have been critically surveyed by Blau (1946). He begins with the observation that preferred laterality is a developmental acquisition of man. This is shown by phylogenetic evolution from *neutral ambi-laterality in animals and early man* to the general preference for *dextrality* in homo sapiens. Ontogenetic development repeats the same progress of maturation from infantile bilaterality to preferential dextrality in about 85 - 90% of all humans. Reflecting neuro-muscular maturation, neutral ambilaterality is commonly found among peoples with " primitive " cultures, notably the Bushmen and Hottentots. Blau further writes that this undeveloped state is common among mentally retarded individuals, increasing in direct proportion to the decrease in intelligence.

Frequently, *uniform* dextral dominance of the right side concerns hand, leg, eye, possibly also the ear, and the cranial and oral muscles. Multiformity appears as *mixed laterality,* such as in dominance of the right hand and left eye. Mixed dominance of various effector and receptor organs must not be confused with ambilaterality, as defined above. In adults, *bilateral activity* may occur in three forms: (1) *Ambidexterity,* the equal and similar dexterity or adroitness in both hands; (2) *Neutral ambilaterality* without preferential dexterity in any hand; and (3) *Ambisinistrality* or ambilevousness, the equal absence of dexterity in both hands, making them equally sinistrous or maladroit (doubly left-handed).

(b) *Evolution of human dextrality* can be traced back at least to the birth of toolmaking in the Stone Age more than 100,000 years ago. Blau states further: " Recent careful archaeological research shows quite definitely that hand preference among aboriginal man in the Stone Age, as among animals, was equally divided." Quoting many anthropological references, he concludes the *original existence of about 50% of left-handers.* The same tendency to ambilaterality was found by Mead (1930) among New Guinea children.

This primordial condition was definitely altered during the Bronze Age, when a marked shift toward right-handedness accompanied the evolution of implements. With the growth of culture, dextrality became incorporated into the social code, or as Blau puts it, " Right sided dominance became not only a matter of convenience and necessity but one of religious and magical significance, moral and ethical values." Hence, the polar dualism of words for " right " and " left " in all languages which express the correlation of " right " with positive values, and the opposite for all " sinister " connotations.

It is superfluous to ponder on any priority of either cerebral or manual laterality because this cannot be answered any better than the question whether the chicken or the hen came first. Both hatched from eggs laid by previous hens and so it goes back along the phylogenetic tree to the first fusion of crystals into living protoplasm. In the same manner, one hand began to predominate when the contralateral cortex was ready for expanded manual control, and the brain began to differentiate its specifically human functions when the hands were ready for manipulation.

(c) *Ontogenetic development of laterality* reveals definite similarities with generic evolution insofar as it goes parallel with the child's acquisition of skills in both hands. Usually, at 9 months of age individual learning of preferred handedness begins to be guided by the inborn tendency to unilateral preference. The first attempts at speaking are made at the same age. Between 1½ and 2 years of age preferred laterality begins to show signs of definite establishment. This is the age when post-natal neuromuscular maturation reaches a first peak, and when speech development sets in. Around the age of four years, laterality is clearly formed in most children, just as language development is then sufficiently advanced for simple conversation.

It is a significant fact that girls acquire dextrality more rapidly than boys, just as the female sex is ahead of the male in learning all language functions beyond 1½ years of age. Conversely, the minimal incidence of left-handedness is about 4% for females and 7% for males (Blau, 1946), or approximately twice as many sinistral males as females. The sex difference in developmental disorders of language and speech is even higher, namely by the ratio of about 1 : 4 between female and male. A relatively greater incidence of sinistrality is definitely found among mental defectives, delinquents, and many psychopathic abnormals (Blau, 1946).

Gesell's books (1940, 1943, 1946) offer additional evidence for the inseparable correlation of language and laterality. *Ambilaterals* tend to be retarded in language development and are also likely to show other irregularities in psycho-motor development which usually disappear when clear unilateral dominance has been achieved. Moreover, laterality is closely linked to spatial orientation, or *directionality*. With normally developed orientation, vertical lines are drawn downward, and horizontal lines from left to right in the typical dextrad direction.

(d) *Aphasiology* stands on firm ground, when confirming the empirical observation that aphasic language disorders result from lesions of the *language areas in the left hemisphere* of truly right-handed individuals, and vice versa. Blau (1946) agrees with most neurologists in linking preferred laterality and the faculty of language with the dominant side of the brain. " Both language and the skilled activities of dominance may be regarded as special forms of communication ". Exceptions are due either to erroneous or incomplete determination of lateral dominance, or to faulty interpretation of the causative brain lesion. Further discrepancies may be explained by the complicated problems of *bicerebrality* in left-handers, constitutional *ambilaterality*, pathologic *sinistrality* or *shifted* hand preference (Subïrana, Logos, 1961).

(e) *Dysphemia* or stuttering, that is blocking or repetitive speech, and its relation to problems of dominance has been discussed by numerous authors (see Arnold, 1960, and Luchsinger and Arnold, 1959). In fashion around the turn of our century, this theory has contributed to the modern trend in American education which encourages left-handers to write with the left hand. Although this national experiment has been criticized by other authorities, our present left-handers are still to be seen in a constant struggle with the right-handed script of our civilization. Kainz and others have stressed the fact that true left-handed writing would have to employ

leftward and inverted mirror-writing, such as Leonardo da Vinci developed for the secret script of his personal notes. Consequently, the *dextrad* writing-ductus (towards the right) of *sinistral* writing (with the left hand) is as unnatural as would be *sinistrad* mirror-writing with a *dexterous* (normally adroit) right hand.

Yet, the psycho-neurotic symptom of stuttering has not diminished among American children since the introduction of tolerated sinistrality. Statistical analysis of the problem by many authors (see Arnold, 1960) has disproved the assumed causal relationship between stuttering and left-handedness. Occasional observations of the appearance of stuttering in connection with a forced shift of handedness seem to be due to the educational pressures exerted on the child by impatient parents. In such cases, stuttering stems from a nervous reaction to the child's emotional tensions created by the constant admonitions and prohibitions. I do not believe that brain function, *per se,* could be adversely influenced by even abrupt changes in manual preferences.

(f) *Dyslalia* as a symptomatic group of various developmental articulatory disorders has often been associated with left-handedness. Since the term, dyslalia, describes the symptom of many different disorders of articulation, one should first define what associations are meant. We are obviously dealing with different pathological problems when discussing the articulatory difficulties encountered among the following diagnostic groups : mentally retarded, hypacusic, dysphasic, dysarthric, dysglossic, emotionally disturbed, autistic, or congenitally disposed children, such as those suffering from genetically determined language disability.

It is therefore not surprising to find contradictory evidence regarding the incidence of left handedness among dyslalic children. In a recent study, Buckle (1951) found no difference in laterality between 100 dyslalic and 100 normal children. This tends to disprove the opinion of those earlier observers who noted a frequent association of dyslalia with sinistrality. Again, the conflict of opinions seems to be a problem of semantics. If we define what type of disturbed articulation is being considered, we may arrive at truly significant figures. It seems that the only genetic relationship which may be expected is that of *disturbed dominance and the specific articulatory dyspraxia* of congenital language disability, especially when this is combined with the dysgnostic elements of specific reading disability.

Blau (1946) stresses correctly that the mirror-like inversals of " mirror-speech " or of the " spoonerisms " occurring with cluttered speech arise from a problem in orientation which is a question of *order or spacing in relation to time*. Although such temporal relations are not directly related to the spatial problem of right and left, these inversals involve a similar principle. In fact, the incomplete orientation with regard to both space and time appears to be the fundamental problem of cluttering and all its preceding delays in the development of language.

(g) *Dyslexia and dysgraphia,* as the specific type of reading and writing disability, are now interpreted as but one phase in the child's struggle with his congenital language disability. Many authors agree on the typical combination of this *specific dyslexia with mixed dominance*. Children with specific reading disability exhibit three peculiar traits : a marked tendency to make *reversal errors* in dextrad reading,

a facility in *mirror-writing,* and an unusual fluency in *mirror-reading* of sinistral material. Such mirrored copy is produced by showing normal texts in a mirror, or by typing a mirrored carbon copy on the back of the original with the carbon paper face up. In Blau's opinion, all these visual errors are due to faulty eye sensation resulting from reversed eye movements. Just as manual sinistrality prefers sinistrad writing, left-eyedness leads to the mirror-like inversion of reading and writing towards the left.

Reliable evidence for these correlations of directionality and laterality has been elaborated by gemellological studies. It is an established fact that homozygotic (or monovular) twins often represent a *mirror-like duplication* of the same individual. Thus, one of the identical twins often reflects the right half of what his partner contains as the left image. This mirror-like reflection frequently involves many details of physical asymmetries, such as of the face, eyes, nose, teeth, hair growth, posture, etc. Similarly, one partner is often right-handed, while the other is left-handed. For this reason, such mirror-twins also may differ in their visual-gnostic and grapho-motor abilities, because the sinistral partner may be inclined towards sinistrad mirror-reading and mirror-writing. Multiple homozygotic births may reveal the same proportions. In the case of the Canadian quintuplets, two were dextral, two sinistral, and the median sister (Clara Roman) ambilateral.

Other authors are equally firm in rejecting a causal relationship between specific dyslexia and disturbed dominance. Although emphasizing the preponderance of males among dyslectic patients, Park (1952) opposes the concepts of heredity and mixed dominance. Instead, he assumes a factor of stress with concomitant biochemical changes. Environmental factors, such as insecurity, neglect, emotional imbalance, negativism, etc., are also emphasized.

In his scholarly monograph on the " Master Hand ", Blau (1946) collected and analyzed a great wealth of information on the problems of dominance. His interpretation rejects both concepts of hereditary influences and of congenital language disorders, such as congenital aphasia. According to his deductions, both sinistrality and language disorders are caused by emotional imbalance. He concludes that their relationship is mainly coincidental and linked through their common parentage from the basic psychogenic disturbance. Considering sinistrality a neurotic trait, he condemns the tolerance of sinistral writing habits and finds the alleged dangers of shifted handedness non-existent. This call for a return to uniform dextral education is the important conclusion from this work.

(h) *Disorders of handwriting* are closely related to the basic organization of manual dominance. Many authors have noticed the general tendency in the right-handed to write to the right, and the left-handed to the left. This is best demonstrated by the spontaneous drawings of children. Since the entire motor pattern of *dextrad and dextral writing* (with a dexterous right hand) in Western culture is built on the numerical prevalence of dextral dominance, every imitation of Western script by sinistral writing (with a dexterous left hand) is fundamentally unnatural. If sinistrals were to develop their innate dominance for all acts of manual preference without

coercion, then they would have to be permitted to use mirror-writing, the only true inversion of the abducting Western script. Since this cannot be done, *sinistral but dextrad* writing by adducting movements of the left hand remains an inadequate solution of inverted laterality.

When examining the problem of disturbed writing from an evolutionary viewpoint, one must consider the archaeological evidence in its proper perspective. Evolution of writing was greatly influenced by the invention of writing implements and the surfaces to be written on. Blau (1946) reminds us that the form of letters and the direction of the writing duct were determined by the surface and the writing tool : " Stone writing seems to have favoured leftward Semitic writing ; the brush, vertical oriental writing ; and the pen, rightward European writing ". Thus, the Western alphabets originated from the sinistrad Phoenician, later went through a transitional period of alternating writing and reading in both directions, called *boustrophedon,* until the Greeks settled on the dextrad direction of present Western graphic.

This evolution reflects the modern and mature *orientation by dextral laterality,* or as Blau puts it, " language has definite directional pre-requisites which are related to the dominant right hand, the rightward direction of eye-gaze and the dominant eye, all focussing on the language symbol ". Just as the abducting movement of the right hand in Western writing is normal for the right-hander, the abducting movement of the right eye from left to right is normal for reading with dextral eye dominance. Conversely, left-eyedness favours the reversal of mirror-writing and reading.

In conclusion of this tentative evalution of the dominance theory for the explanation of developmental disorders of language, it would seem that its true significance for language research lies in the precise definition of all related concepts and their relationship. Clarification of this problem will require carefully planned studies of all dominance factors in properly diagnosed cases of language disorders of congenital, constitutional, and acquired origin. In addition to such studies of pathological deficits, we need anthropological investigations of the specifically human trait of lateral dominance.

Evolution of language and laterality

Many attempts have already been made to draw conclusions from relics of tools and works of art, on the handedness of the originator and his associates. Often, it has been thought that such relics demonstrate a fairly even distribution of dextral and sinistral products, suggesting a lack of preferred laterality in those epochs.

Other graphic testimonies indicate clearly the artists' awareness of preferential handedness. Stein (1949) shows a reproduction of Bushman Cave Art (after Sollas), representing a cattle raid by 10 distant and 11 larger human figures. All of the larger group carry the weapons with the right hand, except the last and largest figure in the lower right-hand corner which is clearly left-handed (spear-left, shield-right). This would indicate a ratio of 10% sinistrality, exactly as found in modern

civilization (Luchsinger and Arnold, 1959). Since the actions are directed towards the left side, the artist seems to have been a right-hander.

Ontogenetically, the development of language and the establishment of preferred laterality go " hand in hand " at the same pace. The same close relationship is demonstrated by pathological observations. Hence, we are safe in assuming a similar interdependence in phylogenetic evolution of language and laterality since the dawn of cerebral dominance and unimanual preference. Far-reaching conclusions on the evolutionary stage of these two human functions in specific cultural epochs can be drawn by studying the cultural remnants of these periods.

It seems, therefore, necessary to re-evaluate archaeological findings of objects and drawings from specified epochs with the following questions in mind : (1) For which hand were *tools and implements designed* and what is the proportion of specific laterality ? This question answers the distribution of preferential handedness and permits conclusions on the state of cerebral dominance at the time these tools were made. (2) To which side are *faces, figures, and actions oriented* ? This answer gives information on the artist's own handedness, for right-handers tend to orient their drawings towards the left. Again, the numerical proportion of these directions indicates the distribution of preferential laterality. (3) In what hand are *weapons, tools, and insignia carried* ? The proportion of such preferences suggests cultural connotations, such as ethical and magical correlations of dextrality with righteousness. (4) *What objects are depicted* ? If an artist could abstract observations into graphic concepts, he must have known analogous language symbols. (5) *What actions are represented* ? If an observer could graphically define spatial and temporal relations of events, he must have possessed a structured communication system for the differentiation of verbs, nouns, and adjectives. One more word about the methodological approach to this problem. Reproductions of all such relics must be prepared with great care, lest the sides be reversed during photographic processing. Since this mistake has occurred, it is advisable to label clearly each side of the object before taking the first picture.

It should be obvious that more information on the evolution of language and laterality can be gained from the logical analysis of cultural testimony than from intuitive contemplation of anatomic parts. The brain speaks, and not the tongue or the jawbone. Just as the large majority of speech and language disorders have nothing to do with the size, shape, and motility of the tongue, and just as the frequent delays in ontogenetic speech development have nothing to do with the mythical superstition of a " tongue tie ", the phylogenetic evolution of language can never be deduced from the comparative morphology of the peripheral expressive mechanism alone.

Cranial bone fragments provide clues for the reconstruction of configuration and size of the skull. Their shape also permits deductions concerning the size and shape of muscles, such as the tongue, palate, external cervical muscles, etc. However, the motility and the complex patterns of co-ordinated movements of these osseo-muscular mechanisms depend entirely on the organization of the pertinent cerebral innervation.

How much caution is needed for such concatinated deductions—from structure to function and thence to performance—on the establishment of speech-adequate mechanisms is illustrated by Stein's conclusions (1949) from the " Piltdown fossils ".

Human language is a specific function of the human brain. Cranial endocasts give some information on the size and gross structure of the brain. Since, however, we cannot expect ever to find a preserved brain of early man, we can only look for the evidence of his brain function : his specifically human behaviour. Hence, any reconstruction of language evolution is inseparably linked to the understanding of cultural evolution. *When man was ready for culture, he sensed the need for contact (Revesz, 1946) and began to formulate symbolic language.*

Subnormal musical ability

Having observed a large number of clutterers, we have become more and more convinced that a major clue to the problem of tachyphemia is to be found in the *modality of auditory perception.* There is general agreement that congenital language disability and well developed musical talent represent opposite poles of the greatly varied language talent in the general population.

(a) *Antithetic polarity of technical and musical ability.* On the one side may be placed the extreme of predominantly scientific and technical abilities as found in members of materialistically inclined professions such as engineering, chemistry, or physics. Many representatives of these occupations demonstrate a tendency towards a concrete, precise, clipped, or even rigid formulation of thoughts. Their scientific journals are written in a plain, factual and sometimes popular or simplified style which has no space for metaphorical embellishment. Secretaries who type such manuscripts comment on the unusual number of errors in spelling and grammar. More often than not careless articulation and awkward selection of words detract from the intelligibility of their oral expression.

The other extreme is represented by the language mastery of those with humanistic, philological, and artistic talents. In this group belong members of the teaching, philosophical, theological or legal professions. Scientific journals devoted to these fields present masterpieces of refined English prose. Many physicians demonstrate similar inclinations. It is not without reason that the medical schools are integrated in the universities and not with the institutes of technology where many sciences basic to medicine, such as physics and chemistry, are studied and taught.

As a next link in our reasoning, we find in many surveys that musical ability is generally more frequent among members of the humanistic professions. In particular, among persons who have hobbies or abilities aside from those directly related to their vocations, physicians as a group are known for their musical talent. Hence, we arrive at a close general *relationship between language and musical talent.* Although human abilities are distributed in most varied combinations, it is nevertheless a fact that musical inclinations are found most frequently amongst the professions which rely on language skill. On the other hand, technical specialists often exhibit a marked lack of musical interest.

In that sense, we may enlarge Luchsinger's (1959) well-known thesis of the hereditary *language debility type* to include the criteria of refined auditory discrimination. This concept leads us to differentiate the extreme prototype of non-musical language debility in association with unilateral mathematical-technical proficiency from the other extreme of musical and language talent amongst people with predominantly humanistic-philological inclinations. According to our observations, clutterers tend to prefer technical or commercial occupations, where they are not handicapped by poorly refined speech. Most of our cluttering patients are completely uninterested in or even disparagingly opposed to music. Most of them also reveal marked signs of tone deafness or other forms of auditory dysgnosia. A pilot study has proved this observation to be true almost without exception.

When dealing with the problem of musical capacity, it is good to remember that this ability is composed of two main categories : expressive and receptive. The *expressive* elements include all motor actions of singing, whistling, and instrumental *performance*. The receptive elements consist in perception, recognition, and emotional experience of musical form and content. These two categories may be present in various combinations and degrees of development. Conversely, cerebral lesions may disturb them singly or in various combinations (amusia, paramusia).

(b) *Musical Ability and Language.* Many authors have drawn attention to certain basic relations between musical ability and all aspects of human communication. The following conclusions are beyond any doubt : (1) Most of the habitual (or functional) disorders of voice and speech (except stuttering) occur much more frequently in persons with sub-normal musical ability ; (2) This correlation of low musical capacity with disability of oral expression is most marked in all conditions which depend on the feedback mechanism of auditory monitoring (dyslalia, all habitual types of dysphonia, etc.) ; (3) The congenital lack of musical talent is most striking in clutterers, especially when the other elements of specific language disability are prominent. Let us examine next how these clinical observations can be related to the concept of central language imbalance (Weiss, 1950) as the basis of all tachyphemic difficulties.

There is substantial evidence for some sort of leading function of the non-dominant hemisphere in the non-verbal aspects of communication : music, mathematical ability, and abstract reasoning. Pfeifer (1936) demonstrated the larger cortical representation of auditory function in the non-dominant temporal lobe which often contains two transverse convolutions of Heschl, whereas the dominant side shows only one such auditory gyrus. Moreover, Heschl's area is much larger in musical individuals than it is in persons without musical talent. Fry (1957) seems to have a similar concept in mind, when stating that the non-dominant side specializes in the recognition and reception of speech, while formulation and expression of language pertain to the dominant hemisphere.

Analyzing traumatic brain injuries, Poetzl (1947) and Arnold (1951) found a close relationship between brain dominance and symptomatology of cerebral auditory disorders : in all cases of centrally disturbed auditory function, lesions of the thalamus

or of the cortical auditory projection were localized in the non-dominant side. It was concluded that lesions of the highest auditory centres in the *non-dominant side* tend to produce the "highest level" disorders of *paracusis or paramusia* (receptive dysmusia), while leaving all language functions intact.

Thus, we arrive at the following analogy of two types of dysgnostic disorders. Lesions of the non-dominant auditory cortex are associated with the general class of auditory agnosia for non-verbal perceptions whose best-known representative is *receptive dysmusia;* lesions of the temporal auditory areas in the dominant hemisphere produce the general class of symbolic and semantic agnosias with the prototypes of *receptive or nominal dysphasia.*

In this sense, the clutterer's diffuse alteration of brain function seems to affect both hemispheres, producing his dysphasic *expressive paraphrasia* in the dominant side, and the dysmusic *receptive dysgnosia* in the non-dominant side. A further element of *receptive dysphasia,* namely the logorrhoeic overflow of disjointed syntax particles, has a counterpart in the rambling propulsion of the clutterer's *tachylalia.* This parallel points to a defective inhibitor mechanism in the feed-back control which is normally exerted by the temporal lobe. If the problem of disturbed dominance is added to these dysgnostic, dyspractic, and dysphasic features, the clutterer's well-known confusion by the musical elements of speech (pitch, rhythm, and temporal succession) becomes apparent with perplexing empathy.

(c) *Diminished Auditory Discrimination.* This concept of diminished auditory ability among clutterers fits well into the new theory of cluttered speech. At the present stage of general understanding, our attention is attracted by three characteristic features in the clutterer's reduced auditory orientation. These are : (1) A tendency towards abstract thinking with diminished attention to the discrete definition of verbal thought concepts ; (2) A deficient feed-back mechanism of auditory monitoring ; and (3) Poor auditory memory for all language symbols, his own as well as those of other speakers.

The tendency towards abstract thinking appears corroborated by Weiss's observation that in many clutterers thinking proceeds in general images without precise formulation in words. This seems to be similar to any normal person's occasional experience when thinking of a specific acquaintance without being able to recall his name. These patients often admit the feeling that their thought process is completed before the appropriate words have been selected. The time interval between the appearance of the impulse to speak and the recall of the necessary words produces the characteristic repetitions in the clutterer's impatient speech. In that sense, the clutterer's tendency towards thinking in abstract images or relationships prevents him from making the required effort to prepare each utterance by the prior formulating process of inner language or by thinking in concrete language concepts.

Deficient auditory feed-back: When he begins to hear his poorly prepared pronouncement, the clutterer's innate limitation of auditory feedback causes a further handicap by inefficient homeostatic control of his motor speech act. Therefore, he becomes more confused by what he hears himself say which results in the well-known

stumbling over syllables with the various inversions of sounds, syllables, words and sentence particles. Expressed in other terms, the clutterer's auditory observation of his speech is extremely feeble, a fact which seems to be related to a lack of sharply defined ideational images in any sensory modality.

Poor auditory memory has been observed by numerous investigators. Recently, Landolt and Luchsinger (1954) emphasized the clutterer's poor auditory memory for words he has just pronounced. The after-image of his own words fades away more rapidly than in normal persons. For this reason, the clutterer often repeats the same error without noticing it. If the error is brought to his attention, he cannot reconstruct what he has just said. *Objective evidence* for the various elements of this auditory dysgnosia has been found by tests of several psycho-acoustic functions : (1) Differential limen for frequency discrimination, (2) Delayed side-tone effect, and (3) Auditory memory span.

Language and localization

Problems of cerebral localization are especially complicated when attempts are made to corrrelate *specific functions* of communication with certain cortical areas. While it is generally accepted that expressive language formulation pertains primarily to Broca's area and that impressive language perception is based in Wernicke's area, both in the dominant side, the further categories of symbolic, semantic, and syntactic communication cannot be ascribed to limited centres. The dilemma created by the controversy between localizing and holistic aphasiology is particularly confusing with regard to the highest levels of human communication : *graphic language and music.*

(a) *Primary structure and secondary function :* Recently, a clearer understanding has been reached by the systematic application of the fundamental dichotomy between primary neuro-muscular organization and secondary or accessory specialization of function. This dichotomy of *life preserving* and *superimposed activity* governs all functions which serve for communication. The respiratory organs have acquired the secondary function of phonation ; in the alimentary oral portal, the accessory function of articulation has been superimposed ; to digital manipulation of food, tools, and weapons has been added the specialized skill of writing ; auditory recognition of global signals important for the preservation and propagation of life (food, dangers, mates) has been differentiated into the analytical understanding of abstract language symbols ; and lastly, as the latest phylogenetic acquisition, visual perception of attractive and perilous situations (food, mates—stronger animals, fire) has evolved into abstract comprehension of graphic symbols by drawing, writing, and reading.

The anatomical features of *neuronal structure* represent the generic and genetically determined brain organization of each species. Normal humans possess the same basic brain organization with the same type and number of convolutions. What differs individually is the relative size and development of certain functional areas, the musically gifted having a relatively wider cortical representation of auditory function, the talented sportsman being distinguished by excellent cerebellar and extrapyramidal co-

ordination, and so on. Cats or monkeys excel by superb motor agility because their equilibrial and co-ordinative capacities are most highly developed. Thus, brain organization is a result of phylogenesis.

Accessory specialization of function has utilized pre-existing structures for the development of superimposed abilities. It reflects the generic and ontogenetic processes of learning, it enables the individual to develop his structural capacities to various levels of achievement. All human beings have equally organized hands, but only a certain number possess the *cerebral differentiation of talent* which enables them to learn painting or piano playing ; all men have the same brain anatomy, but only those learn to speak, write, and read whose environment offers these skills and who can hear and see enough to perceive them. Hence, specialized function reflects the interplay of generic evolution and individual learning.

(b) *Absolute and relative localization :* The same thesis was elaborated in Leischner's recent monograph (1957) on agraphia and alexia. Beginning with the origin and evolution of graphic skills, he presents a conciliatory interpretation of relative localization. He distinguishes primary and secondary brain functions. *Primary* cerebral functions pertain to the original organization of the brain and are carried out by specific effector organs. For example, the pre-central area for manual activity controls all movements of the hand. These primary cerebral functions are *absolutely localized* in discrete cerebral areas.

Secondary brain functions developed at a stage when the organization of the human brain was essentially completed. They are not carried out by any specific brain areas reserved solely for their performance, nor do they possess their own effector organs. These secondary functions borrow cortical territories and peripheral effector organs which best serve their purposes. Therefore, their central correlation with certain cortical areas is of a loose order. It represents a *relative localization* which is individually variable with regard to place, intensity, and ideational type. The motoric, auditory, and visual imagery types influence the development of secondary functions to a much greater extent than the primary brain functions. Another difference between primary and secondary brain functions is the fact that the latter constitute links in a chain of integrated functional patterns.

Although it is accepted that agraphia and alexia are correlated with defects in the parieto-occipital area of the dominant hemisphere, Leischner prefers to ascribe to this region only the *ability to deal with symbols*. What is known for certain is the fact that this ability is most easily disturbed by lesions of this area. Conversely, the development of secondary brain functions depends on favourable environmental and educational factors.

Typical secondary functions are reading and writing which illustrate the relationship between primary and secondary functions of the brain. The nearer a certain brain function is to the brain's primary activity, the easier it is to localize this function in discrete cortical areas. Contrariwise, the higher a secondary function developed from the primary ones, the more brain areas will it require, and the more diffuse

are its topical relations. In this sense, all aspects of expressive and receptive communication represent secondary or accessory brain functions which should be interpreted in the light of their true relationship to the basic organization of the brain.

Language disability as an example of deficient homeostasis

Having reviewed the manifold psycho-somatic and psycho-acoustic disabilities which produce the cluttering syndrome of language disability, we arrive at the dynamic analysis of the underlying organic deviations in brain function.

(a) *Spatial Organization of the Brain.* The gross anatomy of the brain suggests a condensation of the ramified problems of all communication disorders into two main categories : one, expressive-motor disorders of language formulation, and two, impressive-gnostic limitations of receptive communication. There is no doubt about the topographic separation of these functional categories by Rolando's central fissure which defines the border between the frontal and parietal lobes of each hemisphere.

In the category of active *motoric expression* with pre-central localization, we find the paraphrasic elements of cluttering with the related problems of dysphasic, dysgrammatic, dyslalic, expressive-dysmusic, dysrhythmic, dyspractic, and dysgraphic disorders of psycho-motor co-ordination.

The other order of passive *sensory reception* in the post-central convolutions is involved in the group of dysgnostic, dysmnestic, receptive-dysmusic, and dyslectic limitation of psycho-sensory integration with the environment.

As if presenting a connecting link between these two polar categories of dysfunction, developmental *disorders of lateral dominance* contain both the elements of disoriented motor co-ordination and the problems of confused sensory perception. Hence, the persisting interest of clinical psychologists in the laterality problems which are so obviously associated with the specific disabilities of expressive-motor writing and receptive-sensory reading.

Cerebral dominance is important in another respect. We believe that a limited degree of association between certain brain areas, as integrating centres for specific communicative functions, and corresponding pathologic deficits cannot be eliminated. This view is in no conflict with modern methods of aphasiologic language research which rightfully stress the holistic concept of brain function as an integrated entity. The two hemispheres do not compete with one another for some ambitious predominance. Rather, they supplement their individual functions by constant mutual correlation via the commissural association fibres. Thus, the two *verbal* functions of symbolic (1) reception and (2) formulation of language in the dominant fronto-temporal area, and the two *non-verbal* processes of musical (1) reception and (2) expression in the corresponding region of the non-dominant side are geared like a well-bred team of horses for the same common goal : *semantic and emotional communication with the world around us.*

Following this synthetic interpretation of the cluttering syndrome further, a final link is required to unite the functional anomalies within the two basic categories of (1) pre-central *motor* and (2) post-central *perceptive* types, as well as within the

bitemporal parallelism of (3) dominant *verbal* and (4) non-dominant *non-verbal* communication. Confronted with the " crux " of the two main brain fissures, namely the longitudinal intercerebral fissure and the transverse central fissure of Rolando with the continuing Sylvian fissure, we need a unifying factor to explain the interrelated deviations in these *four prototypes of communication disorders* and their spatial correlation with the *four quadrants of the brain :* (1) pre-central expressive dysphasia, and (2) post-central receptive dysphasia, both in relation to the dominant side, as well as (3) pre-central expressive dysmusia, and (4) post-central receptive dysgnosia, both in relation to the non-dominant side.

(b) *Expression Monitored by Reception.* Now, what could be this unifying principle ? Most likely, it is represented by the *homeostatic feed-back mechanism* with its constant interplay of activating and inhibitory impulses. And where may this homeostatic mechanism be found ? Most likely in close anatomical and physiological relationship with the *activating system in the reticular formation* of the sympathetic network. Let us recall what was said in the beginning : the clutterer is not aware of his disability, but when he is activated to better performance, and when he is alerted to his deficient auditory feed-back, he promptly speaks as well as anyone else—at least for a short while until his activation and his feed-back fade away again. Conversely, temporary cluttering can be elicited in normal persons when subjected to the delayed side-tone experiment (Meyer - Eppler and Luchsinger, 1955).

To test a new hypothesis, clinicians usually resort to animal experiments. Unfortunately, this method would be of no value to our problems, for animals have no language in the human sense and what musical ability the higher mammals may possess, would have no bearing on the specifically human acquisition of musical performance and experience. It is possible that comparative musicology may have something to say about certain parallels in the evolution of language and music. If language had to progress from the stereotyped patterns of animal cries to the highly differentiated symbols of human morphemes, music had to follow an analogous course from the generically fixed trills of bird song to the free invention of form and content by the composing genius. Again, invention, expression and reception of musical structure depend on the highly differentiated cerebral capacities of auditory imagination and discrimination.

After our pilot study had progressed to this point, we found a short report by Wolf and Wolf (1959). These authors present a theory for explaining certain speech disorders, such as stuttering and aphasia, as being caused by an abnormal feed-back mechanism. Quoting previous experiments with delayed side-tone presentation, they conclude that a *delay in the auditory feed-back path produces a special type of stuttering* (cluttering ?), and that ordinary stuttering is not caused by an auditory dead time lag. Consequently, the dead time lag responsible for pathological stuttering must occur in some other part of the speech system. These delay causing factors can occur in anyone. Therefore, normal speakers can and do experience occasional non-fluencies or stutter-like blocks when under some emotional tension. Such stumbling over long or complicated words represents a physiological form of cluttering, for

careful enunciation or simple repetition of the mis-articulated word will promptly correct the tongue slip.

From this view-point the puzzling complexity of the tachyphemic language disability suddenly appears much more simple. The clutterer is primarily disturbed by his conceptual and *paraphrasic* disability of incompletely developed language function; this limitation is complicated by his equally basic psycho-motor *dyspraxia;* and he becomes confused by the deficient feed-back resulting from his sensory *dysgnosia.* Looking at it genetically, and keeping in mind that hearing is phylo-genetically much older than speech, the circle is completed by saying that *the original dysgnostic limitation of perceptual function delays the normal acquisition of language, and the resulting paraphrasic disorders remain uncorrected by the persisting dysgnosia.*

This, then, is our new theory which actually rests on the observations and interpretations of all previous contributors to the riddle of cluttering, particularly those of Freund (1934, 1952), Luchsinger and Landolt (1951, 1955) and Weiss (1935, 1936, 1950). It is now presented to information theory with the expectation that a model analogue of deficient homeostasis will be constructed and the mathematical theory of inefficient feed-back mechanisms will be formulated. Similar suggestions have already been made for speech disorders in general (Arnold, 1960; West, 1957). Considering the large number of unrecognized clutterers, of children labouring with reading problems, and of secondary tachyphemic stutterers—all together affecting 10 to 20 million in the U.S.A.—this new theory for the global handling of all language disabilities should become more universally valid than the well-meant but incompletely understood efforts in the direction of sinistral culture and education.

(c) *Tri-dimensional Systems of Communication, Homeostasis and Space.* Although cluttering is a very frequent speech disorder, it has received little attention in modern speech research. Yet the numerically few studies devoted to the intriguing problems of tachyphemia have demonstrated that it is the prototype of a congenital and familial language disorder. Moreover, cluttering encompasses all degrees of faulty language function from the occasional physiological tongue slips of the distraught normal speaker to the continuously unintelligible speech of the severely language-handicapped clutterer. Hence, cluttering represents a symptomatically enlarged and pathologically modified *model for the study of all normal language functions,* especially from the new viewpoints of information theory and communication research.

Groping for suitable approaches to systematic analysis, one first encounters a tri-dimensional system which seems to bring some order into the perplexing multitude of causes, effects, and their associations. (1) A vertical dimension opens the view into an interrelated group of *etiological factors* which affect the development of psychomotor praxis, sensory gnosis, and symbolic abstraction. (2) On a horizontal scale are displayed the interdependent clinical *types of language disorder* with parallel limitation of the expressive and receptive components of communication in all modalities of motoric, phonic, graphic, rhythmic, auditory, and visual performance. (3) The third dimension of depth demonstrates the many *causal interrelations* among the etiological factors and their pathological consequences.

This spatial orientation in the attempted analysis of disturbed language function has a visible counterpart in the *spatial organization of the brain* in four quadrants. Divided by the transverse direction of the central fissure, (1) the pre-central portion is identical with motoric-expressive actions, while (2) the post-central portion is the seat of sensory-receptive functions. These two halves are sub-divided in the median plane by the intercerebral fissure into (3) a dominant and (4) a non-dominant hemisphere with their correlates of verbal and non-verbal forms of communication. Hence, the fundamental importance of cerebral laterality for all language research.

Following the spatial concept further, the three dimensions of *space, time, and causality* serve for the integrated interpretation of normal as well as disturbed language function and its pathological prototype of cluttered speech. Just as every phoneme contains at least three dimensions of formant spectrum, vocal melody, and temporal patterns of rhythm, cluttered speech deviates likewise in at least three categories of reduced oral articulation, monotonous vocal melody, and distorted respiratory rhythm. The same is true for all other deviations in the clutterer's somato-motor and psychosensory performance, as has been discussed.

Summary

Since the three new concepts of (1) language as a servo-system, of (2) homeostatic maintenance by various cerebral feed-back mechanisms, and particularly of (3) artificial cluttering during delayed side-tone presentation are sufficiently established, the new theory of *language disorders as a result of deficient homeostasis* appears to be ready for mathematical formulation by communication research.

Etiological causes and pathological types of language disorder can be analyzed according to certain well-defined categories. These categories should lend themselves for the construction of analogue models. Applying the specified variables to the performance of such models, one should be able to define the precise modes of the most probable feed-back mechanisms. The outline for such programmes of analogue studies is suggested by our findings.

References

Albrecht, W. (1931). Ueber Konstitutionsprobleme in der Pathogenese der Hals-, Nasen- und Ohrenkrankheiten. *Z. Hals- usw. Heilk.*, **29**, 18.

Albrecht, W. (1940). Erbbiologie und Erbpathologie des Ohres und der oberen Luftwege. In *Handbuch Erbbiologie des Menschen* (Berlin), **4**, 1.

Arnold, G. E. (1951). Untersuchung von zentralen Hörstörungen mit neuen Hörprüfungsmethoden. *Archiv Ohrenheilk.*, **157**, 521.

Arnold, G. E. (1960). Studies in tachyphemia : I. Present concepts of etiologic factors. *Logos (Bull. Nat. Hosp. Speech Dis.)*, **3**, 25.

Arnold, G. E. (1961). Phylogenetic evolution and ontogenetic development of language. *Logos*, **4**, (in the press).

Arnold, G. E. (1961). Phoniatric contributions to language research. *Folia Phoniat.*, **13**, 121.

Blau, A. (1946). The Master Hand. Amer. Orthopsychiat. Assoc., Research Monograph No. 5 (New York).

Buckle, D. F. (1951). Speech defect and lateral dominance. *J. Austral. Coll. Speech Ther.*, **2**, 13.

Carhart, R. (1938). Evolution of the speech mechanism. *Quart. J. Speech*, **24**, 557.

Eustis, R. S. (1947). The primary etiology of the specific language disabilities. *J. Pediatrics*, **31**, 448.

Eustis R. S. (1947). Specific reading disability. *New Engl. J. Med.*, 237, 243 - 9.

Eustis, R. S. (1947). Specific reading disability. *Indep. School Bull.*, April, May.

Freund, H. (1934). Zur Frage der Beziehungen zwischen Stottern und Poltern. *Mschr. Ohrenheilk.*, **68**, 1446.

Freund, H. (1952). Studies in the interrelationship between stuttering and cluttering. *Folia Phoniat.*, **4**, 146.

Fry, D. B. (1957). Speech and language. *J. Laryngol. Otol.*, **71**, 434.

Gerard, R. W. (1959). Brains and behaviour. *Human Biol.*, **31**, 14.

Gesell, A. *et al.* (1940). The First Five Years of Life (New York).

Gesell, A. and Ilg, F. L. (1943). Infant and child in the culture of today (New York).

Gesell, A. and Ilg, F. L. (1946). The Child from Five to Ten (New York).

Kainz, F. (1954 - 56). Psychologie der Sprache (4 vols.) (Stuttgart).

Karlin, I. (1958). Speech and language-handicapped children. *Amer. Med. Assoc. J. Dis. Children*, **95**, 370.

Landolt, H. and Luchsinger, R. (1954). Poltersprache, Stottern und chronische organische Psychosyndrome. *Dtsch. Med. Wschr.*, **79**, 1012.

Leischner, A. (1957). Die Stoerungen der Schriftsprache (Stuttgart).

Luchsinger, R. (1959). Die Vererbung von Sprach- und Stimmstoerungen. *Folia Phoniat.*, **11**, 7.

Luchsinger, R. and Arnold, G. E. (1959). Lehrbuch der Stimm- und Sprachheilkunde, Ed. II (Vienna).

Luchsinger, R. and Landolt, H. (1951). Elektroenzephalographische Untersuchungen bei Stotterern mit und ohne Poltererkomponente. *Folia Phoniat.*, **3.**, 135.

Luchsinger, R. and Landolt, H. (1955). Ueber das Poltern, das sogenannte " Stottern mit Polterkomponente " und deren Beziehung zu den Aphasien. *Folia Phoniat.*, **7**, 12.

Mead, M. (1930). Growing up in New Guinea (New York).

Meyer-Eppler, W. and Luchsinger, R. (1955). Boebachtungen bei der verzoegerten Rueckkopplung der Sprache (Lee-Effekt). *Folia Phoniat.*, **7**, 87.

Orton, S. T. (1937). Reading, Writing, and Speech Problems in Children (New York).

Park, G. E. (1952). Nurture and/or nature cause reading difficulties ? *Archives Pediatrics*, **76**, 401.

Pfeifer, R. A. (1936). Pathologie der Hoerstrahlung und der kortikalen Hoersphaere. *Hdb. Neurol. von Bumke und Foerster* (Berlin), **6**, 533.

Poetzl, O. and Arnold, G. E. (1947). Zentrale Hörstörung. *Mschr. Ohrenheilk.*, **81**, 590.

Revesz, G. (1946). Ursprung und Vorgeschichte der Sprache (Bern).

Seeman, M. (1959). Sprachstoerungen bei Kindern (Halle, Germany).

Stein, L. (1949). The Infancy of Speech and the Speech of Infancy (London).

Travis, L. E. (Ed.) (1957). Handbook of Speech Pathology (New York).

Weiss, D. (1935). Beitraege zur Frage des Polterns und seiner Kombination mit Stottern. *Mitteil. Sprach- u. Stimmheilk.*, **1**, No. 4 - 5.

Weiss, D. (1936). Das Poltern und seine Behandlung. *Mschr. Ohrenheilk.*, **70**, 341.

Weiss, D. (1950). Der Zusammenhang zwischen Poltern und Stottern (Ein Grundlegunsversuch des Stotterproblems). *Folia Phoniat.*, **2**, 252.

West, R. (1957). The neurophysiology of speech. Chapter III in *Handbook of Speech Pathology*, ed. L. E. Travis (New York), p. 72.

Wood, N. E. (1959). Language Disorders in Children (Nat. Soc. Crippled Children, Chicago).

Wolf, A. A. and Wolf, E. G. (1959). Feed-back processes in the theory of certain speech disorders. *Speech Path. Ther.*, **2**, 48.

EFFECT OF SUCCEEDING VOWEL ON CONSONANT RECOGNITION IN NOISE

VICTOR SADLER
University College, London

Eighteen subjects were required to identify 23 consonants each presented in CV pairs with each of five vowels in white noise at a S/N ratio of 6 db. Overall articulation scores varied with the vowel used, and there was evidence of strong interaction between vowels and consonants. Judgements were also influenced by linguistic frequencies. It was concluded that the discriminability of CV letter-names could be greatly improved by manipulation of the vowel.

INTRODUCTION

The primary purpose of this experiment was to suggest brief but distinctive names for the 23 consonants of the Esperanto alphabet. At present all consonant names are formed by adding one and the same vowel, namely /ɔ/—e.g. /pɔ/, /tɔ/, etc. A similar situation exists in certain national languages, for example Turkish (which adds /e/) and Bengali (adding /ɔ/). In English, too, consonants which are similar with regard to articulation often bear similar names (e.g. /em, en/ ; /bi, di/). This increases the chances of confusion whenever listening conditions are not ideal. In present practice, this difficulty is usually overcome by using highly redundant letter-names of the ALPHA BRAVO CHARLEY DELTA type. Long strings of such names become cumbersome, however, and it is of some interest to discover to what extent the required redundancy may be attained in monosyllabic names, where the allocation of vowels to consonants is governed by considerations of discriminability. The broad features of the results are presented here in the belief that their relevance extends beyond the language for which the experiment was conceived.

METHOD

The investigations were in the nature of a pilot study, and were restricted to a single type of letter-name, that in which the consonant precedes the vowel. Alternative systems (VC, VCV, etc.) may be explored at a later date. Each of the five vowels /ɑ, ɛ, i, ɔ, u/ was paired with each of the 23 remaining letters of the alphabet and the series randomized. The list was then recorded on magnetic tape at a speed of 3¾ i.p.s. White noise was simultaneously added at a level subsequently found to be on average 5·6 db. below that of the speaker. The 115 syllables were pronounced with what was felt to be a constant degree of effort, in groups of five separated by approximately 8·0 secs. The temporal separation *within* groups was half this value. Timing was achieved by following light signals from a pre-set electronic switch.

Subjects were eighteen adults, both males and females. All spoke Esperanto, but in the majority of cases English was the native language. They were told that their task

was simply to identify and write down the consonants as they heard them ; also, that each of the 23 consonants would be heard with each of the 5 vowels. The character of the auditorium used made it necessary to resort to a replay-level which introduced a certain amount of distortion ; moreover, there was no control over background noises. The results are therefore those for an extremely noisy channel with some distortion and an upper cut-off at about 6000 cps. It is doubtful whether the lower cut-off is of particular importance.

Results

The results were first cast into a set of five confusion matrices—one for each vowel. These five matrices were then combined, and Table 1 shows the observed confusions summed over the vowel variable. In this table, the columns represent stimulus-categories, and the rows the responses. Below each stimulus-letter is given the equivalent symbol from the International Phonetic Alphabet. The last row of the matrix is the no-response category—where the subject was unable to make a decision. Table 2 was again derived from the original data, this time by summing over the stimulus-consonant: it shows the number of times a given vowel elicited a given response, regardless of whether the response was correct or not. In Table 3, on the other hand, the same cells are occupied by frequencies of *correct* responses. From the ratio of the sums of all entries in Tables 1 & 3 it is seen that the overall percentage of correct decisions was 42·07.

Discussion

Before any conclusions are drawn, the following reservations are called for. First, the subjects used had no previous experience of articulation tests ; consequently they may be open to biases which even a little practice would eliminate. Second, little reliance can be placed on the individual cells of the original matrices, since each consonent-vowel combination was recorded only once, and the possibility of fluctuations in pronunciation is considerable. The fact that presentation was limited to a single order will further tend to bias the responses to certain combinations, especially those at the beginning of the series, where subjects were not yet accustomed to the noise. The effect of these latter peculiarities on the results presented in Tables 1 to 3 will be very much reduced, but should be borne in mind.

Certain tendencies emerge clearly enough in spite of the above reservations. Casual inspection of the response-totals in Table 1 makes it abundantly clear that certain possible judgements are neglected in favour of others, and suggests that the preferences are a function of the relative frequency of the various letters in the language as a whole. These frequencies are known (Sadler, unpublished) and correlate with the response-

TABLE 1

S:	b	c	ĉ	d	f	g	ĝ	h	ĥ	j	ĵ	k	l	m	n	p	r	s	ŝ	t	ŭ	v	z	Total
R	b	ʦ	tʃ	d	f	g	dʒ	h	x	j	ʒ	k	l	m	n	p	r	s	ʃ	t	w	v	z	
b	20	2		3	9	5	1	3			1					1	2	7			1	5	3	63
c		1			1											1		2						5
ĉ			64	1		1	28		1		2	1							19	2		3	2	119
d		5		39	1	6	5	1		3	2					1				1	1	1	12	83
f	2	10	1		9	1		8	9			1				6		3		2		1	1	54
g	1	4		8	1	40	3	3			8					1		2			1	5	6	83
ĝ			2	1			41			2	13								1				1	61
h	2				2	1		4	2							2								14
ĥ									1									1						2
j	1					2	2			56	10											1	1	73
ĵ						1	1				40													42
k	5	7	3	7	15	7		8	31			73				8		10		16	1	1		192
l		1		2	1	1	1	1		3	1		67	5	23	1	12	1		1	11	1	1	134
m					1					2			2	55	7	1					3			71
n				1		1				9			8	10	52						3		1	85
p	12	4	1	2	26		1	11	12		2	3				23		12		12				121
r					1						2		6	6	4		65				13	16	3	116
s		22			2				2			1						22	1					50
ŝ			1				1				2								69					73
t		17	15	2	13		2	10	9			6				10	1	17		41			1	144
ŭ										1			1	2							26	2	1	33
v	2	1			1					6				3							6	44	21	84
z		1									1							3				4	19	28
-	45	15	3	24	7	24	4	41	22	8	6	5	6	9	4	35	10	10		15	24	6	17	340
																								2070

Confusion matrix showing consonant perceived against consonant spoken.

frequencies of Table 1 to the extent of 0·5, by Kendall's rank correlation method ($P < 0{\cdot}001$). It seems, then, that the response-preferences are caused by the linguistic frequencies. An alternative hypothesis which seemed worthy of attention is that both sets of frequencies are related to a common cause—namely the relative intelligibility of the various sounds. It is reasonable enough to suppose that a language might use those sounds most frequently which are most easily perceived. As a measure of relative intelligibility, the totals of correct responses in Table 3 were added as a third set of ranks, and *tau* was calculated for the three possible pairs of variables. Then each of the two hypotheses was tested by partialling out the third variable. Thus the relation between the linguistic frequencies and the "intelligibility" factor, with response-frequencies held constant, amounted only to 0·04, whereas the relation between linguistic and experimental frequencies, with "intelligibility" partialled out, was 0·4.

TABLE 2

R:	b	c	ĉ	d	f	g	ĝ	h	ĥ	j	ĵ	k	l	m	n	p	r	s	ŝ	t	ŭ	v	z	Total
V																								
a	3		18	6	5	16	7	1		18	3	17	17	16	6	2	2	5	14	2	6	12		176
e	13		18	9	1	7	8	1		13	14	17	16	14	12	6	15	4	8	5	8	10	8	207
i			4		2	2	14	1		1	11	6	11	12	15	8	16	6	17	9	2	15	4	156
o	4	1	7	18	1	13	8	1	1	16	8	16	15	12	7	2	17		13	10	5	5	4	184
u			17	6		2	4			8	4	17	8	1	12	5	15	7	17	15	5	2	3	148
Total	20	1	64	39	9	40	41	4	1	56	40	73	67	55	52	23	65	22	69	41	26	44	19	871

Frequency table of consonant perceived against vowel following spoken consonant.

TABLE 3

R:	b	c	ĉ	d	f	g	ĝ	h	ĥ	j	ĵ	k	l	m	n	p	r	s	ŝ	t	ŭ	v	z	–
V																								
a	20		29	16	17	28	7	3		27	3	31	42	18	7	15	2	14	14	7	6	14	3	91
e	24		39	22	4	14	12	2		15	14	40	28	15	23	27	17	13	9	29	8	16	12	31
i	1	2	9	1	18	4	18	4		3	12	17	20	21	29	34	18	10	19	50	4	30	5	85
o	12	2	11	28	10	32	13	2	2	18	9	54	32	12	10	30	38	3	13	33	6	8	4	32
u	6	1	31	16	5	5	11	3		10	4	50	12	5	16	15	41	10	18	25	9	16	4	101

As Table 2, with incorrect judgements omitted.

It seems safe to conclude that the language does *not* make most use of its most intelligible consonants, but that subjects are positively influenced by their statistical knowledge of Esperanto, despite the prior information that all consonants would be heard five times. The McGill (1954) method of analysis indicates that of the estimated response entropy $H'(r)$ of 4·18 bits, the error term $H'_{vc}(r)$ remaining unanalyzed when the two stimulus-variables are taken into account is 1·60 bits. The above correlation with linguistic frequencies probably accounts for the majority of this figure, although it might be expected on other grounds that it should also contain the effects of previous transmission. This latter effect, however, could not be adequately investigated in the present experiment owing to the design limitations.

The only relationship here which may be directly tested by the McGill method is that between the stimulus-vowel and the response frequencies (Table 2), for which the data seem to support the 92 degrees of freedom. Here $T'(v;r) = 0{\cdot}18$, $P < 0{\cdot}001$. The same tendency may be seen in Table 3, which therefore justifies the primary assumption of the experiment, that intelligibility is differentially affected by the following vowel. It is interesting to note that there is also an overall effect of the vowel variable, as confirmed by a χ^2 test between the vowel-totals in Table 3 ($P < 0{\cdot}02$).

Inspection of these totals shows that, in general, consonants are more intelligible where the following vowel has a medium-to-low first formant. There is also a suggestion that recognition may be superior where the second formant is relatively high. The ranking of the vowels does not, on the other hand, appear to be a function of their usual sound-pressure level ranking. This is an observation which finds support in Curry *et al.* (1960) ; their study of English letter-names showed no significant relation between sound-pressure level and intelligibility. They did, however, note a tendency towards grouping of the letter-names by vowels, when ranked by percentage correct, and found that names including /e/ were considerably more intelligible than those formed with /i/—further confirmation of the totals presented in Table 3. The experiments of Miller and Nicely (1955) and of Pickett and Rubenstein (1960) dealt with only one vowel each (/a/ and /ɪ/ respectively). Three vowels (/i/, /a/ and /o/) were used in Pickett's (1958) experiment, and were found to yield strikingly different confusion matrices. No direct comparison of overall articulation scores was made between the vowels, however, and since the experiment dealt with compound consonants, no very useful comparison with the present results can be made.

As a check on the other main finding of the present study—the influence of linguistic frequencies—rank correlations were calculated for the data of Miller and Nicely (1955) and Curry *et al.* (1960), with phoneme and letter frequencies respectively. Unfortunately, Curry *et al.* give no information as to actual response frequencies as distinct from correct responses, and the observed correlation with Dewey's (1923) Table D2 should arise by chance 13 times in a hundred. Miller and Nicely, on the other hand, provide suitable data, and the judgements reported by them correlate well with Voelker's phoneme-frequency rankings (quoted by Wang and Crawford, 1960 ; $P < 0{\cdot}01$). Thus the tendency towards linguistic biases in tests of this kind appears to be general.

Conclusions

1) Succeeding vowels have been shown to influence consonant intelligibility, both as an independent source of variation and also in interaction with the consonant variable. Provisional calculations show that it should be possible, by judicious allocation of vowels to consonants, to produce a monosyllabic (CV) spoken alphabet which would raise the articulation score at a speech-to-noise level of + 6 db. from 42 to 70%. This figure does not take into account the additional improvement resulting from the knowledge that a given vowel can only follow a limited set of consonants, once the letter-names are fixed and memorized.

2) Subjects' response-biases are shown to be influenced by the language from which the materials of the test are drawn. The consonants most frequently attributed to the speaker are those most frequently used in the language as a whole.

References

CURRY, E. T., FAY, T. H., and HUTTON, C. L. (1960). Experimental study of the relative intelligibility of alphabet letters. *J. acoust. Soc. Amer.*, **32**, 1151.

DEWEY, G. (1923). Relative frequencies of English speech sounds (Harvard).

MCGILL, W. J. (1954). Multivariate transmitted information. *Psycho-metrica*, **19**, 97.

MILLER, G. A. and NICELY, P. (1955). Analysis of perceptive confusions among some English consonants. *J. acoust. Soc. Amer.*, **27**, 338.

PICKETT, J. M. (1958). Perception of compound consonants. *Language and Speech*, **1**, 288.

PICKETT, J. M. and RUBENSTEIN, H. (1960). Perception of consonant voicing in noise. *Language and Speech*, **3**, 155.

WANG, W. S-Y. and CRAWFORD, J. (1960). Frequency studies of English consonants. *Language and Speech*, **3**, 131.

ADDRESSES OF CONTRIBUTORS TO VOL. 4, PART 2

ARNOLD, Dr. Godfrey E., 6 East 79th Street, New York 21, N.Y., U.S.A.

FILLENBAUM, Samuel, University of North Carolina, Chapel Hill, North Carolina, U.S.A.

HARMS, Dr. L. S., Department of Speech, Louisiana State University, Baton Rouge 3, La., U.S.A.

JONES, Dr. Lyle V., University of North Carolina, Chapel Hill, North Carolina, U.S.A.

SADLER, Victor, Department of Phonetics, University College London, Gower Street, London W.C.1, England.

WEPMAN, Joseph M., University of Chicago, Chicago 37, Ill., U.S.A.

AN INDEX TO MEASURE CONTINGENCY OF ENGLISH SENTENCES*

SELWYN W. BECKER
University of Chicago

ALEX BAVELAS and MARCIA BRADEN
Stanford University

Several indexes to measure contingency of sentences were constructed by considering nouns, repeated nouns, and total number of words. Contingency was operationally defined as reconstructibility in order to test the several indexes against a criterion. The best form of the index was then selected and retested. The contingency ranking, based on the index, of ten sections of text correlated 0·84 with the reconstructibility ranking. It was concluded that the index is a valid initial approximation to a measure of contingency if contingency is defined as reconstructibility.

Scholars have been concerned with language as a communicative device for many centuries. More recently, part of this interest has been expressed in statistical analyses of language as evidenced in the counting of word frequencies and the calculation of word contingencies. Newman (1951) investigated patterning of vowels and consonants in various languages and found that there were marked differences in the degree to which sequences of vowels and consonants form a restriction on the informational content of the language. He also found the greatest degree of patterning in primitive languages and least patterning in English. Therefore, in information terms, English is noisy and unpredictable. In English there is little structure beyond the 5th or 6th letter (the length of the average word) perhaps, because the statistics of words are on a different level from the statistics of letters. Newman and Gerstman (1952) developed a coefficient of constraint which provides an estimate of the sequential dependencies of letters. They conclude that " neither greater freedom nor greater constraint is discovered when sequences of letters are examined of a length greater than that of one word " ; and they also suggest that " in many respects, analysis at the level of words would be most useful."

Miller, Heise and Lichten (1951) found that a word is harder to understand if it is heard in isolation than if it is heard in a sentence because if one hears the first six words of a sentence, the seventh can more easily be determined for the range of possible continuations is sharply restricted.

This sort of formulation, however, has not been applied to the analysis of sentences. Since the number of possible sentences in English is very large, any estimate of the

* *Support of this study was provided in part by a Public Health Service Postdoctoral Research Fellowship, MF-6846-C, from the National Institutes of Mental Health to the senior author at Stanford University, Stanford, California.*

contingency of any sentence, given a previous one, is probably impossible by any method of exhaustion. It is clear then, that some " short-cut " method is necessary in order to get an approximation of the contingency of sentences. A first attempt at such an approximation is proposed in this paper.

Because all possible sentences are not readily available, it is not feasible to obtain frequency counts of any sentence, so a method of calculating sentence contingencies must be based on either a sampling of sentences or on the components common to all sentences. Then some way of verifying the calculated contingency must be obtained. Thus there is a twofold problem: (a) that of calculating an index of the contingency of sentences and (b) devising a test to determine if the proposed index is a valid one.

Rationale and calculation of the index

A sentence is more than a collection of words because of the structure (grammatical rules) by which the words are assembled and also because the relationship of the words to one another within that structure presumably provides some meaning not given by the sum of each word taken independently. Sentences are also collected within another structure, paragraphs, leading one to assume that there might be some meaning associated with the interaction of all the sentences forming the paragraph that is not given by the sum of the meanings of the separate sentences.

If the word is arbitrarily selected as the basic unit for analysis of the sentence " meanings," what kinds of word must be considered ? Most generally, each sentence contains at least two words defined and denoted within the structure (grammar) as a noun and a verb, with the minor exception of imperative sentences. From an inspection of any given paragraph of English text, one can see at least the following kinds of occurrences: (a) each noun appears only once in a collection of sentences; (b) some nouns appear more than once in a single sentence, but do not appear in the remaining sentences of the collection; (c) some nouns are repeated in different sentences, but not in the same sentences; (d) some nouns are repeated both in the same sentence and in different sentences. The verbs in any given paragraph can also occur in any of the same four patterns. If both nouns and verbs are considered, any combination of the patterns of repetitions and non-repetitions can be observed.

Although other kinds of word (other parts of speech) also appear in sentences, only a noun (either stated or implied) and a verb are necessary for a complete sentence. It was assumed that nouns, as object and concept designates, are the more important carriers of meaning, and therefore the key to sentence contingencies. Based on this assumption, the following formulation is presented.

If nouns appeared randomly, then each noun's appearance would reduce uncertainty by a maximum amount, a condition of total non-contingency. If nouns appeared with some fixed, predetermined order, e.g., one noun repeated indefinitely, no uncertainty would be reduced and there would be maximum contingency or complete predictability given the first word. Thus, an index of contingency should include the ratio of

repeated nouns to total number of nouns. If the assumption that the noun carries the sentence meaning is an imperfect one, then some provision for assessing the value of the remaining words should be included. Although the present attempt is concerned chiefly with nouns, in order to give at least some weight to the frequency with which all other words appear, the ratio of nouns to total number of words is included in the index.

The determination of the total number of nouns in a selection of English text is a relatively simple matter, counting; but for purposes of the index, the number of repeated nouns presents problems. Are nouns repeated in the same sentence to be given the same weight as ones repeated in different sentences? What about words that stand for, or refer to, nouns, i.e., pronouns? It was decided that these questions could best be answered after an examination of some empirical data. To that end, several index values were calculated for the same sections of text. The index took the following general form:

$$\text{Index} = \frac{\text{overlap percent.}}{\text{concept percent.}}$$

The several methods of calculating the index were based on differing definitions of the overlap percentage, (O%), and concept percentage, (C%). Excluding pronouns:

$$O\% = \frac{\text{number of repeated nouns}}{\text{total number of nouns}}$$

and

$$C\% = \frac{\text{total number of nouns} - \text{number of repeated nouns}}{\text{total number of words}}$$

With pronouns still excluded, the O% was calculated in one other manner, and designated "weighted overlap," $\bar{O}\%$:

$$\bar{O}\% = \frac{\text{weighted number of repeated nouns}}{\text{total number of nouns}}$$

where repeated nouns received differential values according to their appearances. If a noun was repeated in the same sentence, it received a value of 1, if it was repeated in a succeeding sentence, a value of 2. Thus, for example, if a noun appears twice in each of two sentences, a value of four would be added to the total number of nouns. A value of four would be added to the weighted number of repeated nouns, one for the second appearance in the first sentence, two for the first appearance in the second sentence and for the second appearance in the second sentence. (If it is desirable to restrict the range of the index to keep it from assuming negative values, change the weights of repeated nouns from 1 and 2 to 1/2 and 1.)

Where $\bar{O}\%$ was calculated, C% was affected since the value given to the number of

repeated nouns changed. Thus:

$$\bar{C}\% = \frac{\text{total number of nouns} - \text{weighted number of repeated nouns}}{\text{total number of words}}$$

Thus, excluding pronouns, two values of the index were calculated:

$$(1)\quad I = \frac{O\%}{C\%}$$

$$(2)\quad \bar{I} = \frac{\bar{O}\%}{\bar{C}\%}$$

Two additional values of the index were calculated as a result of the inclusion of pronouns. Pronouns used in place of a noun which previously appeared in the text were considered noun repetitions (other pronouns were ignored except for their contribution to the total number of words), thus changing the values of O% and C% as well as $\bar{O}\%$ and $\bar{C}\%$:

$$(3)\quad I_p = \frac{O\%}{C\%}$$

$$(4)\quad \bar{I}_p = \frac{\bar{O}\%}{\bar{C}\%}$$

The four methods of calculating the index were subjected to test to determine which of the four was the most efficient measure of contingency.

Procedure

Ten sections, each eight sentences long, of English text were selected from widely varying sources (e.g., Feigl, H. 1959; Inkeles, A. 1959; Pearson, K. 1957).[1] The sections were composed of 150 - 200 words with a mean word length of 183·3 words. Four index values were calculated for each of the ten sections providing four bases upon which the sections could be ranked for contingency.

Contingency was operationally defined as reconstructibility. That is, given a section of text from which sentences are taken and presented in random order, an individual can impose an ordering on the sentences according to some set of instructions. (For instance, " arrange these sentences in the best order," " the order in which you think they should be," " the order which the author used," etc.) In the present instance, each of the eight sentences of a section of text was reproduced on separate strips of paper and presented to a group of subjects with the instructions to " arrange these sentences in the order in which you think they should be." Possible order effects were controlled by rotating the positions of the sentences in the sets presented to the subjects. If a group of subjects exhibited a high degree of agreement on the arrangement of a set of sentences, that set of sentences was considered to have high recon-

[1] *A list of the exact passages used can be obtained from the senior author upon request.*

structibility or high "contingency." Conversely, low agreement on the ranks assigned to the sentences was deemed low contingency. Each subject ordered ten sets of eight sentences.

The subjects were two groups of 26 freshmen and sophomore students enrolled at San Jose State College. The first group's sentence reconstructions were used to determine which of the four forms of the index was most efficient. The degree of agreement of the subjects' orderings of the ten sets of sentences was determined by the use of Kendall's Coefficient of Concordance. The ten sets were then ranked from low reconstructibility (low degree of agreement) to high reconstructibility. Each of the four rankings for contingency, arising from the different forms of the index, was correlated with the reconstructibility ranking, determined from the subjects responses. The form of the index which yielded the ranking that correlated most highly with the reconstructibility ranking was selected as the best, or most efficient, index of contingency. The same sets of sentences were presented to the second group of 26 subjects as a test for the following hypothesis.

There is a positive correlation between sentence contingency, as measured by the Index of Contingency, and sentence reconstructibility.

Results and discussion

A coefficient of concordance, (W), was computed for each of the ten sets of sentences from the responses of the first group of 26 subjects (Siegel, 1956). The ten sets were then ranked for reconstructibility based on the W values and then correlated with the rankings for contingency derived from the use of the four indexes. The various values and the five rankings are summarized in Table 1. The correlations (Spearman Rho) between the rankings are summarized in Table 2.

The index whose ranking showed the highest correlation (Rho = 0·75) with the reconstructibility ranking was $\bar{I}_p$ and on the basis of these results it was decided that $\bar{I}_p = \frac{\bar{O}\%}{\bar{C}\%}$ was the best measure of the contingency of sentences.

The entire analysis, using only $\bar{I}_p = \frac{\bar{O}\%}{\bar{C}\%}$ was repeated with the second group of 26 subjects. The reconstructibility ranking derived from the responses of the second group differed slightly from that of the first group, the correlation between the second reconstructibility ranking and the ranking based on the contingency index reached a value of Rho = 0·64 with $p < 0{\cdot}01$.[2]

The conclusion is made, based on these data, that $\bar{I}_p = \frac{\bar{O}\%}{\bar{C}\%}$ is a valid measure of

[2] *If the rhos are treated as Pearson r's and the standard error of the r's is calculated by using the Fischer z transformation it would be seen that that standard error equals 0·21 while the difference between the correlations is only 0·11 so that there is no basis for considering the two correlations as significantly different.*

TABLE 1

TEXT	W	RANK	I	RANK	$\bar{I}$	RANK	I_p	RANK	$\bar{I}_p$	RANK
A	0.26	6	2.44	8	6.50	8	1.29	1	3.64	6
B	0.24	5	2.69	9	8.88	9	1.60	3	2.07	1
C	0.35	7	0.25	1	0.52	1	2.29	5	15.17	10
D	0.12	1	2.35	7	5.27	7	1.59	2	2.45	2
E	0.36	8.5	1.18	4	1.80	3	2.28	4	9.33	9
F	0.16	2	3.73	10	11.66	10	3.40	9	3.16	4
G	0.36	8.5	0.45	2	0.89	2	2.71	6	6.85	8
H	0.43	10	2.07	6	4.55	6	2.94	7	4.05	7
I	0.23	4	1.61	5	2.38	5	3.06	8	3.21	5
J	0.20	3	0.94	3	2.11	4	4.91	10	2.88	3

Rankings of sets of sentences for reconstructibility and contingency.

TABLE 2

	I	$\bar{I}$	I_p	$\bar{I}_p$
W	0.43*	0.43*	0.40*	0.75**

* $0.05 < p < 0.01$
** $p < 0.01$

Correlations between reconstructibility ranking and contingency rankings.

the contingency of sentences, subject to the restriction that contingency be defined as reconstructibility.

It is, of course, obvious that the index is not a perfect measure of contingency even with the restriction. Further tests with different texts should be made in an attempt to refine the index. When such refinement is achieved, it would then be possible to use the index as an independent variable which, perhaps, would be of greater interest than would be the development of the index for its own sake.

It seems possible that communication between friends and strangers might be characterized by different degrees of sentence contingency. Perhaps the effects of attitude inducing communication varies as contingency varies. More likely than not the decision to continue participation in a communication situation depends on the contingencies of the content of that communication. If one knows what is going to be said at a particular lecture, he is less likely to attend than if he thinks new information will be disseminated.

The formal lecture or speech is a situation where the index might profitably be used as an investigative tool. Questions could be asked regarding fluctuations in the index when lecture material is analyzed in sections of words 1 to 100, words 2 to 101, words 3 to 102, etc.; or sections of words 1 to 100, 101 to 200, 201 to 300, etc. The degree of effectiveness of communication or its interest to the listener could then be associated

with fluctuations in the index. It seems logical to hypothesize that there is a curvilinear relationship between effectiveness of communication and index value. Where the index is very low effectiveness might be low because the concept words would be appearing with great frequency and diversity, approaching a random generation of nouns (concept words); where the index value is very high the constant repetition might induce boredom and lack of interest in the listeners. It should be meaningful then to specify optimum levels of the index in order to enhance communicative effectiveness.

Humour is another possible area of investigation. An old joke usually evokes less laughter than a new one because there is considerably greater contingency associated with the "punch-line."

Does the rate of learning change as the contingency of material to be learned changes ?

Because of these, and many other possible areas of inquiry, it is felt that the index of contingency can be a powerful tool in the investigation of the uses and effects of verbal and written communication.

References

FEIGL, H. (1959). Philosophical embarrassments of psychology. *Amer. Psychologist*, 14, 115.

INKELES, A. (1959). Personality and social structure. In R. K. Merton, L. Broom and L. S. Cottrell, Jr. (Eds.) Sociology Today, 249 - 276.

MILLER, G. A., HEISE, G. A., and LICHTEN, W. (1951). The intelligibility of speech as a function of the context of test materials. *J. exper. Psychol.*, 41, 329.

NEWMAN, E. B. (1951). The pattern of vowels and consonants in various languages. *Amer. J. Psychol.*, 64, 369.

NEWMAN, E. B. and GERSTMAN, L. J. (1952). A new method for analyzing printed English. *J. exper. Psychol.*, 44, 114.

PEARSON, K. (1957). *The grammar of science.* (New York).

SIEGEL, S. (1956). *Nonparametric statistics.* (New York) 229 - 238.

SOME FURTHER RESULTS ON THE RESOLUTION OF AMBIGUITY OF SYNTACTIC FUNCTION BY LINEAR CONTEXT*

ASSYA HUMECKY and ANDREAS KOUTSOUDAS
University of Michigan

A modified procedure for identifying the Russian -o/-e/-ee adverbial modifier or predicative complement and providing its correct English equivalent by mechanical translation.

In an earlier paper,[1] we attempted to formulate an unambiguous set of operational instructions which would enable an electronic computer to deal with a certain problem in the syntax of Russian ; namely, to identify the Russian -o/-e/-ee adverbial modifier or predicative complement, and to supply its correct English equivalent. The set of instructions which we developed was based upon the examination of approximately 31,000 running words of Russian scientific text.

Recently, to test the validity of these instructions, we applied them to a text consisting of 43,212 words,[2] and found that they required only minor modifications to handle the larger sample. On this basis it seems reasonable to suppose that the analysis of more texts is not likely to reveal a need for radical changes in the rules as they now exist.

The following, then, is our modified procedure for dealing with -o/-e/-ee forms.[3]

After the word has been identified as an -o/-e/-ee form, check the following word list and if it occurs in this list, translate accordingly.[4]

(1) *bolee*: Choose 'moreover' if it is immediately followed by *togo* which in turn is not followed by *kotor-* ; choose 'more' elsewhere.

(2) *menee*: Choose 'nevertheless' when it is immediately preceded by *tem ne* ; choose 'less' elsewhere.

* *We wish to thank W. H. Dewey, J. O. Ferrell and L. Matejka of The University of Michigan, and W. Bond of I.B.M. and C. Dawson of Syracuse University for their constructive discussions which were very helpful in the preparation of this paper.*

[1] *Koutsoudas, A. and A. Humecky, Ambiguity of Syntactic Function Resolved by Linear Context,* Word 13:3.405-15 (1957).

[2] *Our sample was chosen from the* Žurnal ėksperimental'noj i teoretičeskoj fiziki (*Journal of Experimental and Theoretical Physics*) 29:5,6.537-692,748-61 (1955).

[3] *Henceforth we shall use the following signs:* (/) = (or) ; (+) = (obligatory) ; (~) = (not, negation) ; (±) = (optional).

[4] *The following is a list of -o/-e/-ee forms the translation of which will always be the single English equivalent indicated:* dalee 'further', interesno 'interesting', izvestno 'is known', kačestvenno 'qualitatively', količestvenno 'quantitatively', krajne 'extremely', ljubezno 'kindly', naibolee 'most', neposredstvenno 'immediately', nezavisimo 'independently', neveliko 'small', nevozmožno 'impossible', otdel'no 'separately', preimuščestvenno 'mainly', primenitel'no 'with respect to', primerno 'approximately', principial'no 'in principle', ranee 'earlier', ravno 'equal', soglasno 'according to', svojstvenno 'proper', veliko 'great', želatel'no 'desirable'.

(3) *neobxodimo*: Choose 'necessarily' if it is immediately followed by *dolžn-*; choose 'necessary' elsewhere and preface it with 'is/are' if *neobxodimo* is neither preceded nor followed by an auxiliary word.[5]

(4) *neskol'ko*: Choose 'somewhat' if it is immediately followed by a verb[5] or by an ACM[6] not in the genitive case; choose 'several' elsewhere.

(5) *otnositel'no*: Choose 'relatively' if it is immediately followed by a participle or an adjective; choose 'with respect to' elsewhere.

(6) *sobstvenno*: Choose 'proper' if it is immediately followed by a noun and reverse the order of *sobstvenno* and the noun in translation; choose 'actually' elsewhere.

(7) *sootvetstvenno*: Choose 'corresponding to' if it is immediately followed by the dative case of a noun or ACM; choose 'respectively' if it is immediately preceded or followed by more than a single noun, formula, or ACM; choose 'correspondingly' elsewhere.

(8) *sovmestno*: Choose 'together' if it is immediately followed by *s*; choose 'simultaneously' elsewhere.

(9) *vozmožno*: Choose 'maximally' if it is followed immediately by an -e/-ee form; choose 'possible' elsewhere and preface it with 'is/are' if *vozmožno* is neither preceded nor followed by an auxiliary word.

If the -o/-e/-ee form is not found in the above list, continue examination of its environment[7] and apply the translation rules below.

(1) Choose the adverbial translation if:

(a) the -o/-e/-ee form is preceded or followed by a verb other than infinitive; is preceded by an auxiliary word *and* followed by an infinitive; or is preceded by an infinitive *and* followed by an auxiliary word.

Examples: temperatura *bystro* vozrosla
the temperature *quickly* rose
sosud oxlaždaetsja *medlenno*
the vessel cools off *slowly*
možno *formal'no* vvesti
can be *formally* introduced

(b) the -o/-e/-ee form is preceded or followed by an ACM; if the ACM preceding the -o/-e/-ee form is a *pronoun*, however, an ACM which is *not a pronoun* must follow the -o/-e/-ee form.

Examples: polučennye *vyše* rezul'taty
the results obtained *above*

[5] *The designation "auxiliary word" refers to the following words only (and their inflected form, if any):* byt', stat', moč', pozvoljat', (o)kazat'sja, prixoditsja, sleduet, načinat', neobximo(st'), vozmožno(st'), možno, nel'zja, dolžno. *The terms "verb" and "infinitive", however, refer to any verbs and infinitives, except the auxiliary words.*

[6] *An ACM (Adjective Class Member) includes adjectives, participles, ordinal numerals, and pronouns. See* "Ambiguity of Syntactic Function Resolved by Linear Context," *op. cit. p. 407.*

[7] *By environment we here mean a maximum of two units on either side of the ambiguous form(s). A unit is any word or punctuation mark except the following: (1) words in brackets, (2) all non -o/-e/-ee adverbs, and (3) the words* a, by, no, libo, ne, and že.

v slučae *medlenno* menjajuščegosja polja
in the case of a *slowly* changing field
za ėto *beskonečno* maloe vremja
in this *infinitely* small time

(c) the -o/-e/-ee form is preceded by a period or comma and is followed by a comma which in turn is not followed by *čem* or *čto*.
Examples: *Dejstvitel'no,* značitel'noe . . .
Indeed, a considerable . . .
svojstva, *verojatno,* smožet
property, *probably,* will be able to

(d) the -o form is preceded by a dash.
Example: *—približenno.*
—approximately.

(e) the -e/-ee form is preceded or followed by the abbreviation *sm.* which in turn is not preceded by a numeral.
Examples: (sm. *niže,* ris. 14)
(see *below,* fig. 14)
(*podrobnee* sm. 2)
(*in more detail* see 2)

(2) If two consecutive -o/-e/-ee forms occur, choose the adverbial translation for both if rules (1) (a) or (1) (b) apply. If neither of these two rules apply, choose the adverbial translation for the first and the adjectival for the second (reverse the order if the first -o/-e/-ee form is a member of our first word list).
Examples: vozbuždenija *dovol'no často* soprovoždajutsja
the disturbances are *rather frequently* accompanied
proisxodit *krajne redko.*
occurs *extremely rarely.*
mogut *dovol'no sil'no* otličat'sja
can *rather strongly* differ
provedeno *soveršenno analogično*
(was) conducted *absolutely analogously*
v predpoloženii *dostatočno sil'no* vyražennogo
considering a *sufficiently strongly* expressed

(3) The remaining -o/-e/-ee forms to which the above rules do not apply are predicative and receive an adjectival translation which is prefaced by 'is/are' in all cases except the following: *obyčno* immediately preceded by *kak* and followed by a comma.

Once the word is established as a predicative complement (either by the word list or the rules), use the following additional rules for the re-ordering of the verb 'to be' and the placement of pronouns, articles, and the conjunction 'than'.

(1) In a sequence + *čem* + -e/-ee + formula/noun, put 'is/are' after -e/-ee and 'the' before it; do not translate *čem*.

Example: , čem bol'še X
, the greater is/are X

(2) In a sequence + *tem* + -e/-ee + comma + *čem*, or + *čem* + -e/-ee + comma + *tem*, put ' is/are ' before -e/-ee ; and translate *tem* and *čem* as ' the '.
Example: značenie X tem bol'še, čem . . .
the value of X is/are the greater the . . .

(3) In a sequence ~ *tem* + -e/-ee + comma + *čem*, put ' is/are ' before -e/-ee ; translate *čem* as ' than '.
Example: značenija X v 2-3 raza bol'še, čem
the values of X is/are 2-3 times greater than

(4) In a sequence + verb + -e/-ee + ACM/formula/noun, put ' than ' after -e/-ee.
Example: veličina ėnergii okazalas' men'še vysoty
the magnitude of energy turned out to be smaller than the degree

(5) In a sequence + period + two -o forms + noun, put ' is/are ' after the two -o forms.
Example: Osobenno xarakterno otnošenie
Especially characteristic is/are the relation

(6) In such sequences as (1) + period + -o + comma + *čto*, (2) + kak + -o/-e/-ee ~ ACM, (3) + period + two -o/-e/-ee forms + infinitive, and (4) ~ auxiliary word + -o/-e/-ee + infinitive, put ' it is ' before -o/-e/-ee ; erase ' is/are '.[8]

Examples: (1) Suščestvenno, čto
It is essential that
(2) kak izvestno
as it is known
(3) Bolee estestvenno sčitat', čto
It is more natural to suppose that
(4) Tak kak trudno predpoložit'
Since it is difficult to imagine

[8] *These rules do not account for the correct English translation of such Russian examples as* " V etom slučae *osobenno* jasno, čto ", ' In this case *it* is especially clear, that '. *Although such phrases did occur in our text, their patterning was not immediately apparent and must, therefore, await an analysis of larger samples of text.*

THE MEASUREMENT OF GRAMMATICAL CONSTRAINTS

H. H. Somers, S.J.
University of Louvain

By means of the coefficient of constraint (D) it is possible to measure the constraint exercised by one grammatical type on another, calculating the redundancy of the nth grammatical type when only the first of the chain is known.—The results of the computation are presented, calculated on three Greek texts of the New Testament: the Gospel of St. John, St. Paul's First Epistle to the Thessalonians and the Epistle to the Hebrews. The conclusion is reached that Paul's Epistles have identical features and that, taking into account the individual differences between Paul and John in level of entropy, the constraint is greatest for the second type in the chain, the relative constraint being practically the same for the three texts.

The problem

Mathematical communication theory has been used by Shannon to calculate the information transmitted by sequences of n letters. These formulas have also been used to calculate the uncertainty of the nth letter when $n - 1$ letters are known. These measures of linguistic constraint are equally adjustable to the measurement of grammatical constraint: that is, the cohesion of words. One of the aspects of this cohesion is the impact of one linguistic type on another in the sentence. When we wish to calculate the strength of this impact, we can try to answer questions like the following: when one knows that the first word is a preposition, how accurate will be the prediction that the next word will be a substantive; when one knows that the first word is an article, what is the probability that the third word will be a verb? The coefficient of constraint, as we shall define it, gives an over-all measure of the mean cohesiveness of the text and is an indication of stylistic versatility, regularity or rigidity in the use of grammatical types, such as substantives, verbs, articles, adjectives, prepositions, etc. The present investigation is a part of a series of studies on the mathematical analysis of language and style with special application to biblical texts (Somers, 1959). The different categories of grammatical types have in human language the role of coded indications of the functions of these concepts in the sentence; specifically, the distinct " indicators " of the " substantiveness " or the " adjectiveness " of a word are very redundantly coded expressions of the relatedness of the words in a sentence. They express a certain degree of cohesion of the signs in human language; this degree we want to measure.

Method

The first step in the analysis of constraint is the coding of the text. Therefore we choose a 1,000-word sequence in each of the three texts, which are as much different from one another as possible : The Gospel of St. John, The Epistle to the Thessalonians and the Epistle to the Hebrews. Each word is coded as a number indicating the category of grammatical type to which it belongs, as follows :

Code

1 substantive
2 verb
3 article
4 pronoun
5 preposition
6 adjective or adverb
7 conjunction (and)
8 particle (except subordinations, conjunctions and negations)
9 subordinations (including relative pronouns)
0 negations

In this way we obtained a long chain of 1,000 symbols for each text.
E.g. (Gospel of St. John) :

5 1 2 3 1 7 3 1 2 5 3 1 7 1 2 3 1 4 2 5 1 5 3 1 6 6 6 2 7 5 4
2 8 6 9 2 5 4 1 2 7 3 1 2 3 1 3 1 7 3 1 5 3 1 2 7 3 1 4 0 2 2
1 2 5 1 1 4 1 4 2 5 1 9 2 5 3 1 9 6 2 5 4 0 2 4 3 1 8 9 2 5 3
1 2 3 1 3 6 9 2 6 1 2 5 3 1 5 3 1 2 7. . . .

From such chains it is easy to compute successively the probability of each symbol and of each digram of two symbols. Further it is possible to obtain different kinds of digrams : first, one symbol and the immediately succeeding which we indicate (1,2) ; secondly, one symbol and the third or the fourth or more generally the *n*th symbol after the first considered, which we indicate as the digrams (1,3), (1,4), (1,5) . . . (1,*n*). Thus we may obtain different tables of transition probabilities. Calculating the conditional entropy of the second member of the digram gives us the solution of our problem: a measure of the indetermination or uncertainty, if we calculate the entropy, a measure of the constraint if we calculate the redundancy of the symbol.

Results

First we obtain a table with the probabilities of the symbols in each chain (Table 1). From Table 1 we obtain the H_1 (entropy) of the chain : (first order entropy), according

TABLE 1

	JO.	THESS.	HEBR.
0	022	026	024
1	213	225	218
2	233	151	183
3	139	138	146
4	131	137	110
5	083	112	095
6	044	051	073
7	069	071	044
8	025	052	057
9	041	037	050
s	1000	1000	1000

Frequency of symbols.

to the formula of Shannon:

$$H = -\sum_i p_i \log_2 p_i \tag{1}$$

	JO.	THESS.	HEBR.
H_1	2.95	3.06	3.06

The first finding is the striking similarity of the two Pauline texts and their dissimilarity to the Gospel of St. John. As we shall see, this fact will be confirmed by all the other results.

Next we obtain the tables of the frequencies of the digrams (1,2), (1,3), (1,4), (1,5), (1,6), (1,7), (1,12). The results for (1,3) and (1,4) of the Gospel of St. John are shown in Tables 2 and 3 ; (the results for (1,4) to (1,12) are practically identical).

From these tables we calculate the entropy, the conditional entropy and the coefficient of constraint according to the following formulae :

$$H(i,j) = -\sum_{i,j} p(i,j) \log_2 p(i,j) \tag{2}$$

$$H_1(n) = H(1,n) - H(1) \tag{3}$$

$$D = \frac{H_1 - H_1(n)}{H_1} = 1 - \frac{H_1(n)}{H_1} \tag{4}$$

where $H(i,j)$ is the entropy of the digram ;

$H_1(n)$ is the conditional entropy of the nth type if the first type of the chain is known ;

$H(1,n)$ is the common entropy of the first type with the nth type ;

$H(1)$ is the entropy of the first order ;

D is the coefficient of constraint.

TABLE 2

	0	1	2	3	4	5	6	7	8	9
0	1	3	3	1	3	2	1	3	0	5
1	4	51	63	14	19	31	3	18	6	3
2	2	49	42	33	41	19	19	12	6	10
3	4	31	41	26	15	3	4	6	3	6
4	5	20	34	20	12	5	2	19	5	9
5	1	22	16	13	11	4	3	6	0	6
6	0	10	10	5	4	6	5	1	2	1
7	2	17	8	18	9	9	5	1	0	0
8	1	5	8	2	6	1	1	0	0	1
9	2	4	8	7	11	2	1	3	3	0

$N = 998$

Gospel of St. John: (1,3) digram frequencies.

TABLE 3

	0	1	2	3	4	5	6	7	8	9
0	1	3	7	3	3	3	0	2	0	0
1	7	53	50	33	28	9	7	11	5	9
2	4	45	54	36	29	24	6	17	7	10
3	3	28	39	18	11	10	8	9	4	9
4	2	27	25	14	25	10	4	14	4	6
5	1	19	15	11	11	4	8	5	2	6
6	1	11	12	7	6	2	4	1	0	0
7	0	17	15	10	6	11	4	5	1	0
8	2	3	6	3	4	3	0	3	1	0
9	1	6	10	3	8	6	3	2	1	1

$N = 997$

Gospel of St. John: (1,4) digram frequencies.

TABLE 4

	JO.	THESS.	HEBR.
$H(1)$	2.95	3.06	3.06
$H_1(2)$	2.56	2.69	2.70
$H_1(3)$	2.82	2.96	2.92
$H_1(4)$	2.89	2.99	2.99
$H_1(5)$	2.87	2.97	2.99
$H_1(6)$	2.87	2.97	2.97
$H_1(7)$	2.88	2.98	2.99
$H_1(12)$	2.88	2.98	2.99

Entropy.

TABLE 5

	JO.	THESS.	HEBR.	MEAN
$D_1(2)$	0.132	0.121	0.118	0.123
$D_1(3)$	0.044	0.033	0.046	0.041
$D_1(4)$	0.020	0.022	0.022	0.021
$D_1(5)$	0.027	0.029	0.022	0.026
$D_1(6)$	0.027	0.029	0.029	0.028
$D_1(7)$	0.024	0.026	0.022	0.024
$D_1(12)$	0.024	0.026	0.022	0.024

Coefficient of constraint.

In Table 4 are presented the results of these computations : the total sequential entropy of the *n*th type if the first is known. In Table 5 we find the coefficient of constraint, which is a relative figure varying from 0.00 to 1.00.

DISCUSSION AND CONCLUSIONS

Two major results are obtained : as shown in the first figure, there is a difference in level between the Pauline writings and the Gospel of St. John together with a striking similarity of the function of constraint. Secondly, as shown in the second figure, there is practically no statistical difference within the Pauline writings even when conceptually so different as the First Epistle to the Thessalonians and the Epistle to the Hebrews.

Third, considering the last figure, the relative constraint is identical in the three chosen texts.

Clearly, the constraint will be greater if we take into account the elements themselves, but the coefficient of constraint is a much more reliable and precise index because it is based on the totality of the given elements and their frequency.

The *D*-coefficient is a very reliable and precise instrument for purposes of exact measurement. A long and painstaking study of each of the grammatical types in all the major texts of the New Testament has given the same results as are presented here and obtained with a direct and simple method.

More generally, Shannon found that the total redundancy of the English language should be about 0.77. Here it has been found that about 15% of the redundancy is due to grammatical (and logical) constraints of the type studied ; so there remains 62% for phonetic, syllabic and stylistic constraints. It is known that about 50% of the total redundancy is due to phonetic and morphological constraints, so that there remains 12% for broader logical and other constraints.

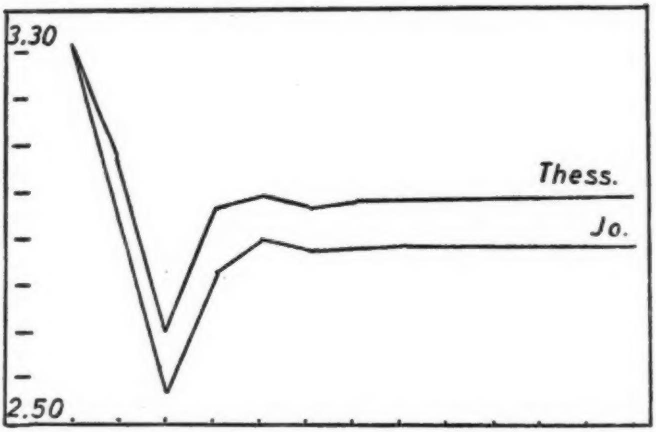


Fig. 1. Conditional entropy for digrams: First Epistle to the Thessalonians and Gospel of St. John.

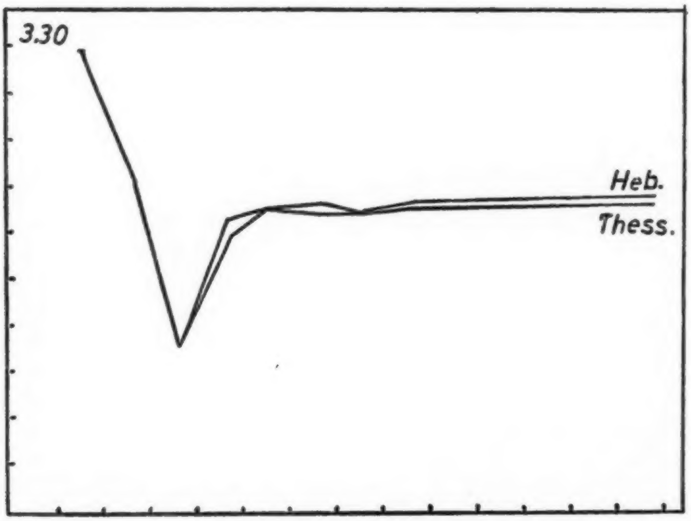


Fig. 2. Conditional entropy for digrams: First Epistle to the Thessalonians and Epistle to the Hebrews.

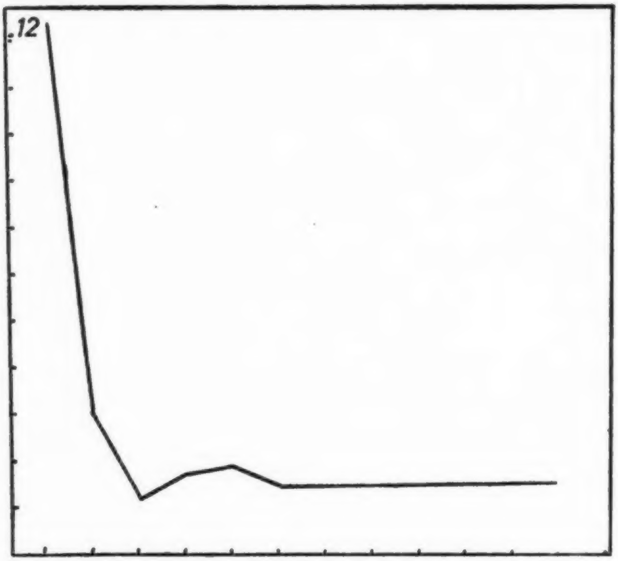


Fig. 3. Mean function of constraint.

References

Shannon, C. E. (1949). The Mathematical Theory of Communication (Urbana).

Somers, H. H. (1959). Analyse mathématique du langage. I. Lois générales et mesures statistiques (Louvain-Paris).

II. Différences individuelles et facteurs psychologiques du style (in the press).

PREDICTION OF WORD-RECOGNITION THRESHOLDS ON THE BASIS OF STIMULUS-PARAMETERS*

KLAUS F. RIEGEL and RUTH M. RIEGEL
University of Michigan

The effect of stimulus-parameters on word-recognition thresholds was investigated. Fifty words were tachistoscopically administered to 24 subjects with an average age of 16·4 years. Of the 40 parameters analyzed, the following were the best predictors of the thresholds and were included in a multiple regression equation: (a) Classification of words into concrete nouns vs. all others; (b) Classification of words with vs. without prefixes; (c) Logarithms of word-frequencies; (d) Number of letters.

The multiple correlation (0·74) and the correlation with the logarithms of word-frequencies (−0·50) are surprisingly low.

The correlations could be raised considerably if counts of children's rather than adult's language were used. Based on this finding and the high correlation between thresholds and our classification of concrete nouns and other words, it was concluded that recognition of words is to a greater degree dependent on the frequency with which subjects had prior experiences with objects (or perceptual images) rather than on the frequency with which subjects had perceived or used the names attached to them.

Despite some initial suggestions by Cattell (1885, 1886) and Ranschburg (1913, 1914), little attention has been given to the analysis of stimulus-parameters which affect the recognition thresholds for verbal material, such as number of letters, number of syllables, word-frequency, etc. Only during the last decade, and as part of the resurgence of interest in social perception, considerable effort has been directed toward such an analysis. Most of the resulting studies deal with visual recognition of syllables or words, employing methods in which either degree of illumination or, more often, tachistoscopic exposure time was increased in equal or logarithmic intervals. Particular emphasis has been given to the effect on thresholds of word-frequency, word-probability, or word-familiarity. In order to control this parameter the following procedures have been applied.

(1) The acquaintance with the stimulus-material has been varied by using nonsense syllables or words in an alien language. In these studies, subjects had to go through a pack of cards and had to read and/or spell the printed syllables before thresholds were measured. Usually, the frequencies of the syllables were varied in steps of:

* *The study was conducted at the Department of Psychology, University of Hamburg (former Director, Dr. C. Bondy) and was, in part, aided by the Foundations' Fund for Research in Psychiatry, New Haven, Connecticut. During parts of the analysis the second author held a Post-Doctoral Fellowship from the National Institute of Mental Health, United States Public Health Service. The computations were completed at the Computing Center of the University of Michigan with the assistance of P. Tomlinson. The authors are indebted to the above persons and institutions but particularly, to the students who helped to collect the data.*

25, 10, 5, 2, and 1. Using this procedure, many investigators (Solomon and Postman, 1952; King-Ellison and Jenkins, 1954; Baker and Feldman, 1956; Postman and Rosenzweig, 1956; and Forrest, 1957) obtained correlations up to $-0{\cdot}99$ between logarithms of word-frequencies and tachistoscopic recognition thresholds.

(2) Miller, Bruner, and Postman (1954) varied word-frequency by using eight-letter sequences of four orders of approximations to English. The higher the order, i.e., the more familiar the sequences to English readers, the lower were the thresholds. As first suggested by Cattell (1885, 1886), the amount of information perceived at a given speed seems approximately constant for subjects.

(3) In most studies such as that of Howes and Solomon (1951), meaningful words have been used as stimuli. These authors measured thresholds for 60 words of different length representing the six value areas proposed by Spranger-Allport-Vernon. Almost all words were abstract nouns, verbs, or adjectives. The correlations between logarithms of word-frequencies as determined by the Thorndike and Lorge counts (1944) and median or mean recognition thresholds ranged from $-0{\cdot}56$ to $-0{\cdot}79$. The authors also observed that subjects often tended to guess words which have higher frequencies of occurrence than the stimulus words. Accordingly, frequency seems to influence subjects' responding rather than perceiving. This generalization is further supported by the work of Goldiamond and Hawkins (1958).

In addition to assessing the effect of word-frequency, Howes and Solomon found a decrease in threshold as a function of practice. Other significant findings indicate that:

" With frequency held constant, (1) letters composed of only a few continuous lines (I, J, L, T) tend to have low duration thresholds; (2) letters in which the white background is chopped up by numerous lines (M, N, S, W), or those in which the white background lies at the centre of the letter so that its lines are closer to surrounding letters than to each other (U), or those whose striking similarity of form leads to conflicting responses (C, G, O), tend to have high thresholds; (3) words with average syllable lengths of less than 2·75 letters (which was the mean for all 60 words) tend to have lower thresholds than words made up of longer syllables; (4) words with repetitive patterns of letters, of the form AA, ABA, ABAB, tend to have low duration thresholds." (Howes and Solomon, 1951, pp. 408-409.)

Kristofferson (1957) reported a substantial correlation between Nobel's index of meaningfulness and thresholds. However, his findings are not congruent with Taylor's (1958) results, since different methodologies were employed. Finally, a number of investigators (Mishkin and Forgays, 1952; Forgays, 1953; Orbach, 1952; Melville, 1957; and Terrace, 1959) have observed that recognition is facilitated if words appear at the right of the fixation point, possibly because attention is immediately directed toward the first letter, which is more important for recognition than the subsequent letters. This result also sheds light on the effect of word length on recognition threshold, since longer words are more likely to extend widely over the visual field.

In spite of the detailed information reported, little is known about the precise interaction of the various factors. The following study was undertaken to derive a multiple regression equation for the prediction of word-recognition thresholds on the basis of

stimulus-parameters. A comprehensive analysis of this type was particularly desirable in our case, since German words and subjects were employed, and all previous studies have been conducted with English-speaking subjects. Since we obtained a moderate correlation only between logarithms of word-frequencies and thresholds, most of our discussion will centre around differences between our own and earlier results and around the attempt to derive interpretations congruent with both.

Methods

The present study is part of an investigation on perception, association, and verbal achievements of young and old subjects. Here, an attempt was made to transform five multiple-choice achievement-tests (synonyms, antonyms, selections, classifications, and analogies: Riegel, 1958, 1959a) into conditions of an experiment on perception. For this purpose, samples of ten stimulus and ten response words representing all levels of difficulty of the test items, were selected from each test and presented tachistoscopically to the subjects. In order to make the experimental results obtained for the five different tests and three different conditions comparable with each other inter-word differences had to be eliminated, i.e., the recognition thresholds had to be corrected for differences in the objective characteristics of the words used. This led to the following analysis, which is based on data for the 50 stimulus words only.

Twenty-four subjects were selected from files of a skilled-trade school in Hamburg. There were twelve boys, most of whom were trained as metal workers, and twelve girls, who were trained as technical draughtsmen and stenographers. The mean age for the total sample was 16·4 and the standard deviation 0·9 years.

The tachistoscope used[1] resembled that of Dodge-Gerbrands. All 50 words (given with their English translations in Table 1) were typewritten in capital letters on 6 × 9 cm. translucent slides leaving one blank space between adjacent letters. The words appeared at the centre of a visual field illuminated in a light reddish colour (due to the particular neon bulb used). The screens were located at a distance of 30 cm. from subjects' eyes. The subject pushed a button to release the stimulus-words; but the experimenter could override and parallel the subject's actions. In order to eliminate fluctuations in accommodation and adaptation, the exposure of the words was immediately followed by the exposure of a blank screen. Subjects were also instructed not to remove their faces from the eye-shade.

As a demonstration, the word "KELCH" (goblet) was administered. The ten stimulus words of each test were randomized for each subject. Seven days elapsed before the next set of ten words was administered, and the order of the five sets was systematically varied between subjects. The following instruction was given:

"When you press that button, a word will appear on the illuminated screen. Your

[1] *The authors are indebted to Dipl. Psych. D. Wendt, Dr. H. W. Wendt, Mr. R. Jacob, Dr. H. Reinecke, and Dipl. Ing. K. Biereichel for the construction of the apparatus. A detailed description of the apparatus and the procedures has been prepared by Wendt, 1959.*

TABLE 1

STIMULI	TRANSLATIONS	TESTS	THRESHOLDS	FREQUENCIES
VOGEL	(bird)	Ag	55	2.49
KLEID	(dress)	Cl	61	2.53
SEE	(lake, sea)	Cl	65	3.32
KOHL	(cabbage)	Cl	69	1.61
LIED	(song)	Ag	69	2.73
KÄSE	(cheese)	Cl	71	1.97
FEUER	(fire)	Se	73	3.56
STEIN	(stone)	Se	73	3.31
ZEITUNG	(newspaper)	Cl	73	2.96
WASSER	(water)	Ag	75	3.69
SCHIFF	(ship, vessel)	Se	75	3.22
STALL	(shed, stable, stall)	Se	75	2.20
GARTEN	(garden)	Se	84	3.07
BENZIN	(gasoline, fuel, gas)	Se	84	1.23
ZEICHNUNG	(drawing, design, sketch)	Cl	87	3.02
BUCHT	(bay, inlet)	Ag	87	1.91
GEWICHT	(weight)	Ag	90	3.09
WÜRFEL	(die, dice, cube)	Se	90	2.37
HAMMER	(hammer, knocker)	Cl	90	2.06
LOKOMOTIVE	(engine)	Ag	94	1.83
FOLGE	(order, series, consequence, continuation, *succession*)	At	98	3.71
SCHMIED	(smith)	At	98	1.93
EINGESTEHEN	(allow, grant, confess)	At	98	1.53
REGEN	(rain)	Cl	98	3.11
KAFFEE	(coffee)	Se	106	2.74
GARBE	(*sheaf*)	Se	106	1.72
ÜPPIG	(rank, *luxuriant, sensual, sumptuous, voluptuous*)	At	106	1.61
POSAUNE	(*trombone*, trumpet)	Cl	106	1.30
KIRCHE	(church)	Ag	110	3.29
BANAL	(*banal, commonplace, trife*)	Sy	116	−1.00
FLIEGEN	(fly)	Cl	125	2.66
ORIGINAL	(original)	At	140	2.56
VERWANDELN	(*alter, figure, mute, transform*)	At	145	2.29
ORCHESTER	(orchestra, band)	Se	145	2.29
FACKELN	(hesitate, delay, *linger, loiter, temporize, waver*)	Sy	145	1.81
SPEICHER	(silo, storehouse, warehouse, *granary*)	Sy	150	1.69
LINKISCH	(awkward)	Sy	150	0.70
MONOTON	(monotonous, *humdrum*)	At	158	0.70
GROTTE	(*grotto*)	Sy	162	1.40
ERINNERN	(remember, remind, recall)	Ag	166	2.80
GESTATTEN	(permit, allow, grant)	At	170	2.72
ERBARMEN	(pity)	Sy	170	2.10
HINDERN	(prevent, hinder, *impede, obstruct*)	At	190	2.81
TRIFTIG	(*cogent, conclusive, convincing, forcible, plausible, valid*)	Sy	194	−1.00

TABLE 1 (cont.)

EIGENSINN	(*obstinacy, selfwill, stubbornness*)	Sy	198	1.86
BELEIDIGEN	(*insult, offend*)	At	202	1.93
OPTIMAL	(*optimal*)	Sy	210	−1.00
BERSCHRÄNKUNG	(restraint, *confinement, constriction, limitation*)	At	210	2.51
MEHRERE	(several, *sundry*)	Ag	234	3.20
RELIKT	(reminder, *relics, residue*)	Sy	424	−1.00

Stimulus-words, their translations[a], name of tests from which words were selected[b], mean recognition thresholds (msec.), and logarithms of word-frequencies.

[a] *Translations printed in italics did not occur in the children's counts.*
[b] Sy = *Synonym Test,*
At = *Antonym Test,*
Se = *Selection Test,*
Cl = *Classification Test,*
Ag = *Analogy Test.*

task is to recognize the word and to tell me after each trial what you have seen. I will always inform you when you may push the button."

All reports were recorded by the experimenter. Subjects had to recognize the stimulus-words correctly three times in succession. Provided that no further trials were requested by the subject, the third from the last measure was regarded as the threshold. The exposure time, which could be varied between 50 and 8800 msec., was increased in equal steps of approximately 10 msec. (or exactly 5 scale-units on the apparatus timer), beginning with the lowest exposure time of 50 msec.

RESULTS

In the first part of the analysis, the recognition thresholds were correlated with approximately 40 word-parameters, the most important of which are given in Table 2. The relative variation of letters (or vowels) denotes the number of different letters (or vowels) divided by total number of letters per word. Both correlations are negative and suggest that the fewer the repetitions, the more rapid is the recognition of the word. This should be particularly true for vowels which, being few in number, carry on the average, less information than the consonants. Thus, if one particular vowel occurs repeatedly in a word, the subject has to study the other elements quite carefully. These results support the findings of Ranschburg (1913, 1914), but not those of Howes and Solomon (1951). They are consistent with our positive correlation on letter repetition and symmetry. The corresponding index was computed by giving a weight of 1 to repetitions of letters, separated by one to three letter-units from each other

TABLE 2

	WORD-PARAMETERS	r	r_{pb}
1.	Number of letters	0.42**	
2.	Number of different letters	0.28*	
3.	Relative variation of letters	−0.25	
4.	Relative variation of vowels	−0.32*	
5.	Letter repetition and symmetry	0.36*	
6.	Letter frequencies	0.39**	
7.	Number of syllables	0.38**	
8.	Prefix (=3) vs. no prefix (=2)		0.44**
9.	Suffix (=3) vs. no suffix (=2)		0.11
10.	Logarithm of frequencies of first syllable	0.48**	
11.	Transitional probabilities	−0.17	
12.	Logarithms of word-frequencies	−0.50**	
13.	Concrete (=2) vs. non-concrete (=3)		0.66**

* ($p < 0.05$) ; ** ($p < 0.01$).

Correlation coefficients between word-parameters and word-recognition thresholds.

(e.g., EGE, NZIN, or NDELN), and by subtracting a weight of 1 if there was a zero-distance between the two letters (e.g., EE). A great number of related measures have been analyzed in our study, but the one above proved itself of greatest value for the predictions.

The index of letter frequencies was operationally defined by the occurrence of particular letters in those words which make up the lower and the upper 27 percentile of the threshold-distribution respectively. The frequencies of occurrence of the letters F, L, and S, in a given word were subtracted from the frequencies of the letters M, N, R, and T and a constant of three was added. The first are those letters which are found more frequently in words that are easy to recognize; the latter in words that are hard to recognize.

Next, the effect of syllables on the recognition thresholds was analyzed. The eighth and ninth correlations suggest that prefixes as well as suffixes delay recognition since they are most likely to increase the length of words. In particular, prefixes transmit a greater amount of information than suffixes and will be of greater importance for the recognition of the whole word whereas suffixes are highly redundant to the preceding parts and may be guessed more easily. Accordingly, subjects have to devote more time to the recognition of the first than of the last parts of words. This result is congruent with findings on the significance of the first letters of words as reported by Sumby and Pollak (1954), Miller and Friedman (1957) and Bruner and O'Dowd (1958). If, however, the first syllable occurs very frequently in the language (as do the very common prefixes GE-, VER-, EIN-, etc., in the German language) the information gained will be relatively small since the word may still continue in many different ways. Accordingly, our tenth correlation reveals that if a rare syllable appears at the beginning of a word it will facilitate recognition.

The above findings are supported by our index of transitional probabilities. This measure was obtained by adding all frequencies of words that contain the same initial letter sequences and then dividing the sum by the number of letters per stimulus word. Starting with the fourth letter, the changes in frequencies from letter to letter differ quite markedly between words. The computations and differences in transitional probabilities are demonstrated in the following examples:

Sequences	*Frequencies*	*Probabilities*
ORCH	104	1·0
ORCHE	104	1·0
ORCHES	104	1·0
ORCHEST	104	1·0
ORCHESTE	104	1·0
ORCHESTER	104	1·0
		$\Sigma = 5{\cdot}0$
ZEIC	3029	
ZEICH	3029	1·0
ZEICHN	1836	0·6
ZEICHNU	476	0·3
ZEICHNUN	476	1·0
ZEICHNUNG	476	1·0
		$\Sigma = 3{\cdot}9$

The frequencies of the words as well as of the first syllables were determined by using Kaeding's count (1898) of 11 million words. This count provides three numbers for each word: (a) Frequency of the word in its simple form; (b) Total frequency of the word, i.e., number of times the word appeared either in simple form or as initial, middle or final part of compound forms; (c) Number of times the word was spelled with a capital initial; (d) We computed the frequencies with which the word appeared in simpler form or as initial part of a compound form. The four correlations did not differ markedly from each other. The correlation on total word-frequencies was the highest and is reported in Table 2. The logarithms of the total word-frequencies are given in Table 1.

The highest correlation with the recognition thresholds was obtained by separating the words which denote concrete objects from those which do not. This classification includes on the one hand all nouns which clearly have perceivable referents and on the other all abstract nouns, adjectives, and verbs. It should be noted that in most earlier studies, such as that by Howes and Solomon (1951), a relatively small number of concrete nouns was employed.

All intercorrelations between the parameters have been computed and a set of four variables was selected to derive a multiple regression equation. The intercorrelations of

TABLE 3

WORD-PARAMETERS	$\hat{Y}$	1.	2.	3.
1. Concrete (=2) vs. non-concrete (=3)	0.66			
2. Prefix (=3) vs. no prefix (=2)	0.44	0.49		
3. Logarithm of word-frequencies	−0.50	−0.37	−0.02	
4. Number of letters	0·42	0·46	0·57	−0·08

Intercorrelations between word-parameters and word-recognition thresholds.

these four parameters are given in Table 3. The following equation and a multiple correlation of 0·74 were obtained:

$$\hat{Y} = 13{\cdot}29\,C + 8{\cdot}04\,P - 4{\cdot}98\,W + 0{\cdot}87\,L - 2{\cdot}08$$

C = concrete vs. other words, i.e., concrete words were given a score of 2, the others a score of 3. P = prefix vs. no prefix, i.e., words with prefixes were given a score of 3, those without a score of 2. W = logarithms of word-frequency to the base ten as determined by using Kaeding's count (1898). When no frequency is reported in the count, a logarithm of −1.00 is assigned to the word. L = number of letters per word. The last index does not contribute significantly to the prediction of recognition thresholds.

Generally, the degree of multiple correlation as well as of all single correlations between the parameters and recognition thresholds is quite disappointing. The high intercorrelations may be responsible for the low degree of multiple correlation, but in particular this result is due to the low correlation between word-frequencies and thresholds which in earlier studies has exceeded the multiple correlation obtained above. Before any conclusions can be drawn, we have to examine some possible determinants of this result. There are at least four arguments which have to be discussed.

THE EFFECT OF WORD-FREQUENCIES ON WORD-RECOGNITION THRESHOLDS

(1) Since the German count by Kaeding (1898) is relatively old and may not adequately represent the language of today, differences in recency of the German and American word counts have to be considered as a possible source of error. There may also be differences between the languages, i.e., one may speculate that the effect of word-frequency on recognition thresholds is different in the two languages. Since these arguments cannot be attacked on the basis of our study and a modern count of the German language is not available, it seems quite meaningless at the present time to pursue these questions any further.

(2) The low correlation between logarithms of word-frequencies and thresholds may be due to the particular shape of the distributions of word-frequencies. We became

aware of this possibility in a pilot analysis of the present data (Riegel, 1959b) which included fewer words with high frequencies. The omission of these words curtails the distribution and reduces the correlation from $-0{\cdot}50$ to $-0{\cdot}05$. Primarily this marked change is due to the assignment of logarithms of $-1{\cdot}00$, i.e., frequencies of 0·1, to stimulus words which did not appear in the word count at all. In this procedure we followed Howes and Solomon (1951), although we realized that the corresponding words would occur only once among 110 million German words and once among 45 million words of the American Magazine Count while the next words recorded occur as frequently as 7 times among the 11 million German words (only words with frequencies of 8 or above are reported) and once among the 4·5 million words of the American Magazine Count.

Aside from these rare words, the distributions of the logarithms are quite regular but somewhat flat. Our logarithms of the word-frequencies are, on the average, slightly higher than Howes and Solomon's (2·10 in comparison to 1·73). Both standard deviations are equal to 1·09 after appropriate multiplications of the frequencies by the approximate ratio of the size of both counts had been made. In conclusion, we may point out the close dependency of the correlations between logarithms of word-frequencies and thresholds on the particular sample of words used. Owing to skewness and kurtosis of the distributions, Howes and Solomon's results, as well as the present findings, have to be regarded as upper limits rather than average estimates of the relationship. However, this does not provide an explanation of the differences between the two studies.

(3) The most difficult problem to discuss concerns differences in the methods applied. We increased the exposure time in equal intervals of approximately 10 msec. or exactly 5 scale units beginning with a duration of 50 msec. In a different procedure from our own, one could increase the exposure time in geometric or logarithmic steps. In most experiments, however, a combination of these methods has been applied. Thus, Howes and Solomon (1951), using no rigid schedule of progression, increased the intervals in steps of 10 or 20 msec. below the exposure time of 200 msec. In the range from 200 to 450 msec. they increased the time in units of 20 or 30 msec., and above 40 msec. they used steps of 40 msec. Moreover, they presented each flash duration twice.

Arguments can be raised in favour of both procedures. Our method can be defended for its rigour which leaves less leeway to the experimenter whereas the method of Howes and Solomon is less time consuming and may better eliminate errors due to practice and fatigue. Although a detailed discussion of this methodological problem has to wait until both methods have been experimentally compared in studies of verbal perception, a few remarks are appropriate on the particular effect of these methods on the obtained correlations.

One may argue that frequent exposures alone will lead to recognition of words even when the duration time is below the threshold level. The argument does not necessarily imply any notion about "subception", since for recognition of compound stimuli, such as words, a subject may well concentrate his efforts at any time on the

identification of particular parts of that compound such as single letters or letter sequences. Using this strategy, a subject is likely to name the word after a few exposures whereas he would not have succeeded if he had tried to recognize the whole word all at once. This strategy will yield its greatest efficiency in the recognition of long and difficult words. Its successful use is revealed by many reports from subjects. In the case of a difficult word, subjects usually begin to report parts of it whereas, in the case of easy items, complete words are given quite frequently as first answers.

Accordingly, the upper end of the scale of thresholds will be curtailed in using our procedure but this effect cannot be as marked when the exposure time is increased in progressively longer steps. Thus, the variation of thresholds will be smaller in our case than for results obtained by the other method and this, in turn, will have its effect on the correlations between logarithms of word-frequencies and thresholds. Comparing the distributions of thresholds as given in Table 1 and as estimated from Fig. 1 of Howes and Solomon's paper does not lead to definite conclusions. Although our standard deviation is smaller than that of Howes and Solomon (63.21 vs. 66.71), this measure is confounded with many other factors, such as the means (125.40 vs. 174.88), the sample of words used, apparatus, size and type of letters, degree of illumination, etc. Thus, the empirical data do not lend themselves to an easy support of our argument. On theoretical grounds, however, we can conclude that the differences in results are at least partly due to differences in the methods applied.

(4) As mentioned above, the sample of words used has an important impact on the results. In comparison to earlier studies, we employed many more words of the common language (particularly concrete nouns) whereas Howes, Solomon, and others used abstract nouns, adjectives, and verbs which are not quite as frequent in everyday language. One of the most striking results of our analysis is the high correlation between thresholds and classification of the words into concrete nouns and those which do not denote physical referents.

It has been recognized in developmental psychology that the acquisition of words by young children proceeds approximately in the order: concrete nouns, verbs, adjectives, abstract nouns. These findings have some additional support from clinical observations of aphasic patients: Ribot's law states that aphasic recovery of language functions proceeds in almost the same order as above. This evidence and the interaction between logarithms of word-frequencies and our classification led us to suspect that higher correlations between logarithms of word-frequencies and thresholds would be obtained if counts on reading material of children rather than adults were to be employed in such an analysis.

The effect of general perceptual experiences on word-recognition thresholds

Objections have been raised by Davids (1956) and Daston (1957) about the use of word counts to estimate facilitation of perceptual processes or probabilities of response-emission. Word counts such as the Lorge Magazine Count may be representative of information the average adult has obtained through his reading but it is questionable

whether the information of college students, the ordinary subjects for experimentation, are greatly dependent on such reading material. It is not so much their present reading habits that must be taken into account but rather those of the years prior to the testing. In the case of college students this would mean a consideration of their earlier school reading.

Unfortunately, for our present analysis, corresponding counts are available for American subjects only (the basic vocabulary of elementary school children by Rinsland (1945) and of kindergarten children by Horn (1928)). We did not hesitate, however, to translate the German words into English since we were only interested in grade by grade comparisons and did not intend to use our results for prediction of thresholds in the way discussed above. This translation was usually possible without ambiguity. For some of the medium and rare words, however, a number of alternatives offered themselves as possible substitutions. All these alternatives as given in Cassell's dictionary were accepted on principle since under such conditions the resulting increase in total frequency of the rare word could only lower the correlations between logarithms of word-frequencies and thresholds. Thus, the correlations could be accepted as fair guesses rather than overestimations of the relationship.

In Fig. 1 four series of correlation coefficients are presented for the kindergarten level (Horn, 1928), grades I to VIII (Rinsland, 1945), and the adult level (Thorndike and Lorge's magazine count, 1944). First, correlations between logarithms of word-frequencies and thresholds were highest and of almost equal magnitude up to the sixth grade ; thereafter they drop very regularly down to $-0{\cdot}30$ at the adult level. Realizing that the data derived from the German adult count revealed a correlation of $-0{\cdot}50$ one may argue for a proportional rise in correlation at all younger age-levels if the corresponding data had been obtained from German sources. Second, there is a consistent drop in correlation between our classification of words into concrete nouns vs. other words and logarithms of word-frequencies from $-0{\cdot}69$ at the first grade to $-0{\cdot}10$ at the adult level. This change indicates most clearly to what extent the language of young children is dominated by the use of concrete nouns and to what extent both predictors (our classification and the word-frequency) become increasingly and independently useful the more one proceeds along the age scale. Third, there is a slight increase in partial correlations between our classification and thresholds, holding logarithms of word-frequencies constant at the different grade levels. The lowest correlations of 0·42 occur at the first and second grade, the highest of 0·66 at the adult level and this reveals again the increase in predicative value of our classification. Fourth, there is a fluctuation between $-0{\cdot}41$ and $-0{\cdot}18$ in partial correlations between logarithms of word-frequencies and thresholds, holding our classification constant at the different age levels. Thus, in spite of the loading of children's vocabulary with concrete nouns, logarithms of word-frequencies are independent but relatively poor predictors of thresholds in comparison with our classification.

Undoubtedly our interpretations would have been much improved if supportive evidence had been derived for English words. Nevertheless, the very regular increase in correlations between our classification and thresholds, holding word-frequency

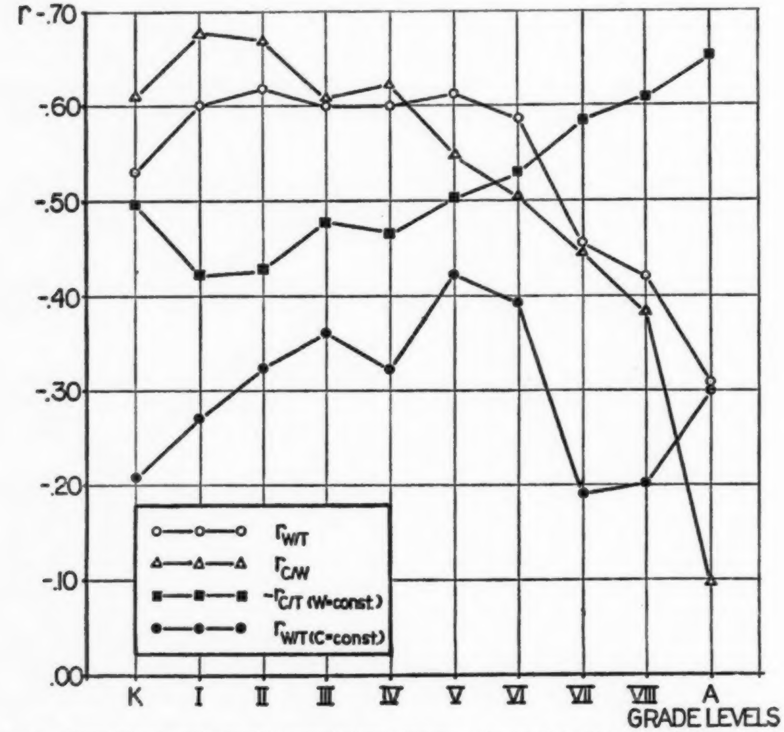


Fig. 1. Correlations and partial correlations between thresholds (*T*), classification (*C*) and word-frequencies for different grade levels (*W*).

constant, seems to warrant the general conclusion that recognition of words is to a higher degree dependent on the frequency with which subjects have had experience with objects (or their perceptual images), rather than on the frequency with which they have perceived or used the names attached to them.

This conclusion opposes earlier interpretations which failed to distinguish between words as mere physical stimuli and as representatives of a more general stimulation evoked by what is called objects and objective processes. Words as physical stimuli are experienced only by speaking, hearing, writing, and reading. Words as more general representatives are experienced during a vast variety of sensory and motor processes. constant seems to warrant the general conclusion that recognition of words is to a Thus, one may read the word " chair " only very seldom ; one may hear the word " chair " more often ; but one is experiencing " chairs " during all of his daily activities. Our findings suggest that frequency of the latter type of experience is also vital for perception of the associated words whereas earlier interpretations emphasized merely the frequency with which we experience words as physical stimuli.

Aside from the correlations based on our distinction of concrete words and those which have no physical referents, there is little evidence to support our conclusion. The findings by Taylor (1958) seem at first to be contradictory. This author presented nonsense syllables together with coloured pictures of familiar objects at varying frequencies to one group of subjects and the syllables alone, but equally frequently, to another group. The results indicated that the frequency with which syllables and objects were associated did not affect the recognition thresholds markedly. One has to realize, however, that the frequency of trials prior to the measurements does not parallel by any means the frequency with which we ordinarily perceive well-known objects. Moreover, the association of nonsense syllables with pictures of familiar objects is most likely to interfere with older verbal habits.

Some indirect support for our conclusion may be derived from the performance of sensorily deprived subjects. In order to satisfy their need for sensory stimulation, these subjects should be likely to engage longer than normals in those sensory-motor activities that remained available to them. McAndrew (1948) has shown that this is particularly true for blind subjects in various modelling tasks. Stated in terms of our interpretation, deprived subjects need a greater accumulation of information through their remaining senses to have the total amount of sensory input correspond to that of normal subjects.

This interpretation concerns only the interdependency of experiences obtained through different sense organs and the overall constancy of total sensory input for different subjects. We may generalize, however, that also sensory-motor experiences related either to symbols or objects are interdependent. Thus, if a person is very well acquainted with the symbol the perception of the object should be facilitated. On the other hand, his familiarity with the object should increase his efficiency in perceiving or using the corresponding symbol. Perception of words is not only dependent on subjects' experiences with these symbols but with the total sensory-motor experience related to corresponding objects or objective processes.

References

Baker, K. E. and Feldman, H. (1956). Threshold-luminance for recognition in relation to frequency of prior exposure. *Amer. J. Psychol.*, **69**, 278.

Bruner, J. S. and O'Dowd, D. (1958). A note on the informativeness of parts of words. *Language and Speech*, **1**, 98.

Cattell, J. Mc.V. (1885). Ueber die Zeit der Erkennung und Benennung von Schriftzeichen, Bildern und Farben. *Philos. Stud.*, **2**, 635.

Cattell, J. Mc.V. (1886). Psychometrische Untersuchungen. *Philos. Stud.*, **3**, 452.

Davids, A. (1956). Personality dispositions, word-frequency, and word-association. *J. Pers.*, **24**, 328.

Daston, P. G. (1957). Perception of idiosyncratically familiar words. *Percept. mot. Skills*, **7**, 3.

Forgays, D. G. (1953). The development of differential word-recognition. *J. exp. Psychol.*, **45**, 165.

Forrest, D. W. (1957). Auditory familiarity as a determinant of visual threshold. *Amer. J. Psychol.*, **70**, 634.

GOLDIAMOND, I. and HAWKINS, W. F. (1958). Vexierversuch: The log relationship between word-frequency and recognition obtained in the absence of stimulus words. *J. exp. Psychol.*, 56, 457.

HORN, E. (1928). The International Kindergarten Union: A study of the vocabulary of children before entering the first grade. (Baltimore.)

HOWES, D. and SOLOMON, R. L. (1951). Visual duration threshold as a function of word-probability. *J. exp. Psychol.*, 41, 401.

KAEDING, F. W. (1898). Häufigkeitswörterbuch der deutschen Sprache. (Berlin.)

KING-ELLISON, PATRICIA and JENKINS, J. J. (1954). The duration threshold of visual recognition as a function of word-frequency. *Amer. J. Psychol.*, 67, 700-703.

KRISTOFFERSON, A. B. (1957). Word recognition, meaningfulness, and familiarity. *Percept. mot. Skills*, 7, 219-220.

MCANDREW, HELTON (1948). Rigidity and isolation: A study of the deaf and the blind. *J. abnorm. soc. Psychol.*, 43, 476-494.

MELVILLE, J. R. (1957). Word-length as a factor in differential recognition. *Amer. J. Psychol.*, 70, 316-318.

MILLER, G. A., BRUNER, J. S. and POSTMAN, L. (1954). Familiarity of letter sequences and tachistoscopic identification. *J. gen. Psychol.*, 50, 129-139.

MILLER, G. A. and FRIEDMAN, E. A. (1957). The reconstruction of mutilated English messages. *Inform. Control*, 1, 38-55.

MISHKIN, M. and FORGAYS, D. G. (1952). Word-recognition as a function of retinal locus. *J. exp. Psychol.*, 43, 43-48.

ORBACH, J. (1952). Retinal locus as a factor in the recognition of visually perceived words. *Amer. J. Psychol.*, 65, 555-562.

POSTMAN, L. and ROSENZWEIG, M. R. (1956). Practice and transfer in visual and auditory recognition of verbal stimuli. *Amer. J. Psychol.*, 69, 209 - 226.

RANSCHBURG, P. (1913, 1914). Ueber die Wechselwirkungen gleichzeitiger Reize im Nervensystem und in der Seele. *Zeit. Psychol. Physiol. Sinnesorg.*, 66, 161 - 248 ; 67, 22 - 144.

RIEGEL, K. F. (1958). Eine Untersuchung über intellektuelle Funktionen älterer Menschen. *Techn. Rep. No. 2* (Psychol. Inst. Univ., Hamburg).

RIEGEL, K. F. (1959a). A study of verbal achievements of older persons. *J. Geront.*, 14, 453.

RIEGEL, K. F. (1959b). Wissenschaftliche und methodische Grundlagen des psychologischen Testverfahrens—Möglichkeiten und Grenzen. (Preisschrift Dtsch. Forschungsgem., Bad Godesberg.)

RINSLAND, H. D. (1945). A basic vocabulary of elementary school children. (New York.)

SOLOMON, R. L. and POSTMAN, L. (1952). Frequency of usage as a determinant of recognition thresholds for words. *J. exp. Psychol.*, 43, 195.

SUMBY, W. H. and POLLACK, I. (1954). Short-time processing of information. *HFORL Rept. TR 54-6.*

TAYLOR, JANET A. (1958). Meaning, frequency, and visual duration threshold. *J. exp. Psychol.*, 55, 329.

TERRACE, H. S. (1959). The effect of retinal locus and attention on the perception of words. *J. exp. Psychol.*, 58, 382.

THORNDIKE, E. L. and LORGE, I. (1944). The teacher's word book of 30,000 words. (New York.)

WENDT, D. (1959). Eine tachistoskopische Untersuchung an älteren Menschen und einer jüngeren Kontrollgruppe mit verbalem Reizmaterial aus dem "Synonym-Test" von Dr. Klaus Riegel. *Dipl. Psychol. Diss., Univer. Hamburg.*

THE SIGNIFICANCE OF CHANGES IN THE RATE OF ARTICULATION

FRIEDA GOLDMAN-EISLER

University College, London

The term "rate of articulation" is applied to the absolute rate of speech, i.e. the rate based on the time of vocal speech utterance exclusive of pauses. The significance of its changes was studied in relation to changes in levels of verbal planning and in degrees of spontaneity. The effect of individual differences was also investigated. While articulation rate proved to be a personality constant of remarkable invariance it also reflects the degree of spontaneity in the production of speech. Variations in level of verbal planning were shown to have no effect on the rate of articulation. The implications of these results are discussed.

INTRODUCTION

A previous investigation (Goldman-Eisler 1956) showed that what is commonly perceived as the speed of talking, or the rate of speech production, is determined by the halts and pauses which interrupt the flow of speech rather than by the speed at which the actual speech movements are performed.

A continuous flow of speech, rarely broken by periods of silence, is felt to be fast speech, and speech the flow of which is halted by frequent pauses of hesitation is experienced as slow speech. The speed of the actual articulation movements producing speech sounds occupies a very small range of variation (4.4 to 5.9 syllables per second were obtained from speech uttered during interviews) while the range of pause time in relation to speech time was five times that of the rate of articulation. Indeed, so rare was a continuous flow of speech in spontaneous utterance in the samples investigated that the variations in speed of talking observed were largely determined by the time spent in hesitation. Naturally, with increased fluency in the utterance of speech the speed of talking becomes a function of the rate of articulation or speed of speech movements themselves.

While pauses had been shown (Goldman-Eisler 1958, 1961) to reflect the process of selection and planning in speech, the comparatively minor variations of the articulation rate have so far remained unexplained. An experiment which was so designed that speech was produced under three different and well-defined conditions may throw some light on the question of the significance of speeding or slowing down the rate of articulation.

Experiment

The experiment which is reported in detail elsewhere (Goldman-Eisler, 1961)) had originally been designed for a different purpose. It consisted in showing, to highly intelligent subjects, cartoon stories without captions (of the kind regularly published in the " New Yorker " magazine) asking them first to describe the content of the stories and then to formulate the meaning, point, or moral of the story. The subjects were also requested to repeat these descriptions as well as the formulations of the meaning (to be referred to as summaries) six times after their first version.

Experimental conditions were thus created for the study of (a) speech produced within a relatively concrete situation, i.e. a given sequence of events (through their description), (b) speech uttered in the process of abstracting and generalising from such events (through summarising their meaning), (c) speech uttered in both these situations for the first time while being planned and organised, being thus new speech, and (d) speech uttered after several repetitions being well organised[1], practised speech. The first two situations (a) and (b) represent two different levels of abstraction in verbal planning, or of redundancy in the coding of information, while situations (c) and (d) represent spontaneous and automatic speech activity respectively.

The speech produced by the subjects was recorded, transcribed and visual records obtained of the sequence of sound and silence. The duration of these was measured, as described before (Goldman-Eisler, 1956, 1958) and the rate of articulation was calculated by dividing the time for the speech sounds only by the number of words produced.

Nine cartoons were described and their meaning summarised by nine subjects.

Results

The measurements were in terms of time (seconds) per word produced and analyses of variance were performed based on this quantity which was normally distributed. The differences compared by analysis of variance (see Table 1) were those between (a) the two types of operation, namely describing events and summarising meaning, (b) between individuals, (c) between cartoons described and summarised, and (d) between the means of the first versions of these verbal formulations and those of their seventh most practised repetitions.

The overall mean time per word produced was 0.268 sec. which gives us the articulation rate of 3.7 words per second.

The mean rates of articulation (in terms of time per word) for description and summaries was 0.269 and 0.267 sec. per word respectively and the variance between these two levels of verbal operation was nil. In other words, the processes of abstraction involved in the formulation of meaning, or of re-coding information in

[1] *The expression " well organised " is used in the Hughlings Jackson sense, meaning that the nervous arrangements for such sequences (speech) are well organised.*

TABLE 1

SOURCE	SUM OF SQUARES	*df*	VARIANCE
Operations			
(Descriptions and Summaries)	0	1	0
Individual Differences	0.15	8	0.019
Cartoons	0.03	8	0.004
Residual	0.78	144	0.005
		161	

F Individual differences / within = 3.80 P = 0.001
F Operations = 0 n.s.
F Cartoons = 0.08 n.s.

Analysis of variance of articulation rates.

TABLE 2

SOURCE	SUM OF SQUARES	*df*	VARIANCE
Individual Differences	0.0199	8	0.00249
First time and after practice	0.0093	2	0.00465
Residuals	0.0179	25	0.00072
		35	

F Individual Differences = 3.4 P = 0.01
F First time and after practice = 6.4 P = 0.01

Analysis of variance, comparing articulation rates of first formulations and after practice.

speech, seem to have no effect whatever on the actual speech movements, while, as shown previously (Goldman-Eisler, 1961) they had a profound effect on the intermittent pauses, leading to a dramatic slowing up of the overall speech rate. Nor did the type of cartoon presented affect the articulation rate.

Individual differences, on the other hand, were highly significant (P = 0.001) thus confirming an earlier suggestion (Goldman-Eisler 1956) that the rate of articulation is a personality constant of remarkable invariance.

This invariance, however, proved subject to modification as a result of practice. An analysis of variance comparing the rates of articulation when formulating speech for the first time and after practice repeating it for the seventh time showed a decrease of time per word from 0.285 to 0.252 sec. for the description, and from 0.277 to 0.245 sec. for the summaries, or an increase of 3.5 to 4.0 words per second for descriptions and 3.6 to 4.1 words per second for summaries. The difference for the total sample was significant at the 0.01 level of probability.

Conclusion and discussion

It may be useful to consider the fact that while the rate of articulation is a constant of such rigidity that it does not respond to changes in the levels of verbal planning, i.e. to distinctly and qualitatively different degrees of abstraction when encoding information into speech, as do the pauses, it does respond to practice.

This corroborates the idea that there is a wider, a more basic difference between speech sequences which are familiar and well learned and those which are spontaneous and organised at the time of utterance, than exists among the spontaneous and newly organised speech sequences which differ in levels of verbal planning, degrees of abstraction involving distinctly different levels of symbolic activity, or in the redundancy of formulation. Such a basic division as the former is, of course, contained in the old-established duality formulated by Hughlings Jackson as voluntary or propositional vs. automatic and inferior speech.

The value of this result for the interpretation of the processes involved in speech production is that it gives us an indicator of unequivocal significance. Its rigidity under the impact of considerable differences in the levels of abstraction on the one hand, and its modifiability with practice, whatever the degree of abstraction or redundancy in coding, on the other, make it an efficient indicator of habit strength, and habit strength only, entering into the production of speech.

In other words, an increase in speed of articulation thus indicates an increase in the use of prepared and well learned sequences, of cut and dried phrases and clichés, of trite and vernacular speech, of commonplace utterances or professional jargon.

It indicates that there is less creative activity and that time serves no function other than that of sound transmission. While the mood which goes with speech that is being organised while being uttered seems to tend towards an arrest of time, that which accompanies speech requiring no further activity beyond the vocalisation of learned connections travels through time at a pace dictated, at the best, only by external requirements such as e.g. intelligibility. Such speech will more easily become subject to corruption in the form of slurring, gabbling, etc., and the reason why we find these characteristics in pathological speech or speech produced under abnormal conditions may be due to the fact that such speech consists mainly of established speech habits.

References

Goldman-Eisler, F. (1956). The determinants of the rate of speech output and their mutual relations. *J. Psychosom. Res.*, **1**, 147.

Goldman-Eisler, F. (1958). Speech production and the predictability of words in context. *Quart. J. exp. Psychol.*, **10**, 96.

Goldman-Eisler, F. (1961). Hesitation and Information. In Proceedings of 4th London Symposium of Information Theory (in the press).

AN EFFECT OF LEARNING ON SPEECH PERCEPTION: THE DISCRIMINATION OF DURATIONS OF SILENCE WITH AND WITHOUT PHONEMIC SIGNIFICANCE*

ALVIN LIBERMAN,** KATHERINE SAFFORD HARRIS, PETER EIMAS,**
LEIGH LISKER,*** and JARVIS BASTIAN****
Haskins Laboratories, New York

Discrimination of an acoustic variable (various durations of silence) was measured when, as part of a synthetic speech pattern, that variable cued a phonemic distinction and when the same variable appeared in a non-speech context. In the speech case the durations of silence separated the two syllables of a synthesized word, causing it to be heard as *rabid* when the intersyllabic silence was of short duration and as *rapid* when it was long. With acoustic differences equal, discrimination proved to be more acute across the /b,p/ phoneme boundary than within either phoneme category. This effect approximated what one would expect on the extreme assumption that the listeners could hear these sounds only as phonemes, and could discriminate no other differences among them; however, the approximation was not so close as for certain other consonant distinctions.

In the case of the non-speech sounds the same durations of silence separated two bursts of noise tailored to match the onset, duration, and offset characteristics of the speech signals. There was, with these stimuli, no appreciable increase in discrimination in the region corresponding to the location of the phoneme boundary. If we assume that the functions obtained with the non-speech patterns represent the basic discriminability of the durations of silence, free of the influence of linguistic training, we may conclude that the discrimination peaks in the speech functions reflect an effect of learning on perception. It was found, too, that discrimination of the non-speech patterns was, in general, poorer than that of the speech. From this we conclude that the effect of learning must have been to increase discrimination across the phoneme boundary; there was no evidence of a reduction in discrimination within the phoneme category.

In studies of the perception of /b,d,g/, /d,t/, and /sl,spl/[1] (Liberman, Harris, Hoffman and Griffith, 1957; Griffith, 1958; Bastian, Eimas and Liberman, 1961; Liberman, Harris, Kinney and Lane, 1961) we have found that discrimination of a

* *This research was supported by the Operational Applications Office of the Air Force Electronic Systems Division in connection with Contract AF 19(604)-2285.*
** *University of Connecticut.*
*** *University of Pennsylvania.*
**** *Now at Center for Advanced Studies in the Behavioral Sciences, Stanford, California.*

[1] *These phonemes were presented in contexts as follows: /b,d,g/ were in absolute initial position before /a/; /d,t/ were in absolute initial position before /o/; /sl, spl/ were in the words* slit *and* split.

given acoustic difference is considerably more acute across phoneme boundaries than in the middle of the phoneme categories. To make the appropriate measurements we had first to identify acoustic variables which are sufficient cues for the perceived phonemic distinctions. For that we were able to fall back on the results of earlier studies (Liberman, Delattre, Cooper and Gerstman, 1954 ; Delattre, Liberman and Cooper, 1955 ; Liberman, Delattre and Cooper, 1958 ; Harris, Hoffman, Liberman, Delattre and Cooper, 1958 ; Bastian, Delattre and Liberman, 1959). Having thus selected for each phonemic distinction an appropriate acoustic cue, we prepared a series of synthetic patterns in which that cue was varied along a single continuum through a range large enough to encompass the phonemes being investigated. To measure discriminability, we arranged the patterns into ABX triads and asked the listeners to decide, on the basis of any similarities or differences they could hear, whether X was identical with A or with B. (A and B were, in fact, always different, and X was always identical with the one or the other.) To find the phoneme boundary, we presented the patterns with instructions to identify each one as /b/, /d/, or /ɡ/ in the first experiment, as /d/ or /t/ in the second, and as /sl/ or /spl/ in the third.

Discrimination was so much better across the phoneme boundary than within the category as to suggest that the listeners could only hear these consonants categorically (i.e., as phonemes), and could discriminate no other differences among them. This suggestion was tested by finding the extent to which the discriminability of the patterns could be predicted from the way in which the listeners had assigned the stimuli to the various phoneme categories. Discrimination functions that were derived on this basis were found to fit rather closely those that had been obtained in the experiments.[2]

From a psychological point of view these results are quite unusual. With stimuli that vary along a single dimension (of frequency, intensity, or duration, for example) one typically finds that subjects discriminate many times more stimuli than they can identify absolutely (Pollack, 1952, 1953 ; Garner, 1953 ; Chapanis and Halsey, 1956 ; Miller, 1956). In everyday experience this is illustrated by the contrast between the ease with which we normally distinguish two tones as being of different pitch and, on the other hand, the great difficulty we have in absolutely specifying the pitch of either one. The very different result in the perception of /b,d,ɡ/, /d,t/, and /sl, spl/ was that discrimination was little better than absolute identification. It is as if our listeners were able to distinguish only as many pitches as they could correctly name.

Viewed from a linguistic standpoint, these results might not appear surprising. Apparently the linguist is prepared to find, with some phonemes at least, that variations in a speech sound will be heard by phonetically naive listeners only when these variations are phonemic. More generally, he might see the extent to which this happens as a precise expression of the degree to which linguistic categories are also categorical in perception.

[2] *In the case of* /b,d,g/ *the fit was better in the Griffith (1958) study than in the experiment by Liberman* et al. *(1957). This was attributed to the fact that Griffith's synthetic speech stimuli were more realistic, and to certain procedural improvements he was able to make.*

Within either a psychological or linguistic framework, the peaks in discrimination should be of interest, we think, because their existence may be an important condition underlying the distinctiveness of speech sounds. Thus, an incoming stimulus which falls ever so slightly to one side of the peak becomes indistinguishable from, and no harder to identify than, a stimulus which lies in the precise centre of the phoneme region. The effect of this should be to reduce the area of uncertainty between phonemes, thereby increasing the accuracy and speed with which the listener sorts the various sounds of speech into the appropriate phoneme bins.

We should note that there appears to be considerable variation among phoneme classes in the size and sharpness of the discrimination peaks, and, correspondingly, in the extent to which the perception is categorical. For some phoneme distinctions there are no discrimination peaks at the phoneme boundaries, and the level of discrimination is far better than would be predicted from the extreme assumption that the listeners can hear the sounds only as phonemes. This kind of result has so far been found in the perception of vowels (Fry, Abramson, Eimas and Liberman, in preparation), and of several prosodic features (tones and vowel duration) which are phonemic (Abramson, 1961 ; Abramson and Bastian, in preparation).

To determine why the discrimination peaks develop at some phoneme boundaries and not at others, we should have to inquire quite deeply into the nature of the perceptual mechanism. For the purposes of this paper it is appropriate only to indicate the broad outline of our hypothesis. We believe that in the course of his long experience with language, a speaker (and listener) learns to connect speech sounds with their appropriate articulations. In time, these articulatory movements and their sensory feedback (or, more likely, the corresponding neurological processes) become part of the perceiving process, mediating between the acoustic stimulus and its ultimate perception. When significant acoustic cues that occupy different positions along a single continuum are produced by essentially discontinuous articulations (as, for example, in the case of second-formant transitions produced for /b/ by a movement of the lips and for /d/ by a movement of the tongue), the perception becomes discontinuous (i.e., categorical), and discrimination peaks develop at the phoneme boundary. When, on the other hand, acoustic cues are produced by movements that vary continuously from one articulatory position to another (as, for example, the frequency positions of first and second formants produced by various vowel articulations), perception tends to change continuously and there are no peaks at the phoneme boundaries. Various aspects of our view have been described elsewhere (Cooper, Delattre, Liberman, Borst and Gerstman, 1952 ; Liberman, Delattre and Cooper, 1952 ; Liberman, 1957 ; Cooper, Liberman, Harris and Grubb, 1961 ; Bastian, Eimas and Liberman, 1961 ; Harris, Bastian and Liberman, 1961), and it will be developed further in papers now in preparation. A theory which is in certain ways related to ours has been put forward in an interesting paper by Ladefoged (1959).

Basic to our speculation about the mechanism which accounts for the discrimination peaks is the simple assumption that they are learned. The primary purpose of the experiments to be reported here is to provide data relevant to that assumption.

Whether the peaks are, in fact, acquired in experience, or whether they are somehow a part of our innately given sensitivity to the acoustic stimuli, is a question of broader scope than is a consideration of any particular mechanism as such.

We cannot dismiss, out of hand, the possibility that the discrimination peaks are innately given. If they are, we should suppose that the earliest speakers of the language wisely chose to locate the phoneme boundaries in the regions of highest discriminability. Assuming, alternatively, that the peaks reflect the learning that has occurred during each listener's long experience with the language, we must then answer a further question concerning the direction the learning has taken. Thus, it is possible that the peak is an increase in discrimination, acquired as a result of the listener's having had for many years to distinguish sounds that lie on opposite sides of the phoneme boundary. Such an effect might prove to be similar to what has been called " acquired distinctiveness " ; for convenience, we will use that term to describe it. The contrary, and equally likely, possibility is that the peak is what remains after discrimination has been reduced by long training in responding identically to sounds that belong in the same phoneme class. This would likely be counted an example of " acquired similarity ". It is also possible, of course, that the observed effect is the sum of both processes: acquired distinctiveness across the boundary and acquired similarity within the category.

As between these two assumptions—that the peaks are part of the listener's innately given sensory equipment, or, alternatively, that they are the result of learning—the latter interpretation is the more likely. One relevant consideration is that languages other than English have apparently located their phonemes differently on the acoustic continua with which we are here concerned. Although there are no data yet available to show that the inflections in the discrimination function are displaced to correspond with the different positions of the phoneme boundaries, the mere fact that the boundaries are differently located is, in itself, presumptive evidence that the highs and lows of the discrimination function are not innately given.

A learning interpretation is also favoured by the fact that the discriminations among synthetic /b,d,g/, /d,t/, and /sl,spl/ were so largely controlled and limited by the phoneme labels. As was pointed out above, this relatively close correspondence between differential sensitivity and absolute identification is in striking contrast to the usual psychophysical result. One may suspect that it has come about because the original or raw discriminations have been radically altered by long experience.

The experiment by Griffith (1958) on /b,d,g/, which we referred to earlier, has provided additional relevant evidence. Using essentially the same second-formant transitions employed by Liberman *et al.* (1957), he added one or another of two constant third-formant transitions which had the effect of changing the positions of the phoneme boundaries. The result was that the peaks and valleys of the discrimination functions shifted accordingly. Though not critical, this evidence strongly supports a learning interpretation.

The experiment with /d,t/ that was referred to earlier was also of a type designed to find out whether the observed peak in discrimination is a result of learning, and,

if so, whether it is a case of acquired distinctiveness, acquired similarity, or both. The point of this kind of experiment is to measure the discriminability of an acoustic variable which cues a phonemic distinction, and then to measure the discriminability of essentially the same variable in a non-speech context. For /d,t/ the acoustic variable was the time of onset of the first formant relative to the second and third. It was found, as has already been pointed out, that discrimination of this variable was better across the phoneme boundary than within the phoneme category. To produce appropriate non-speech controls the experimenters simply inverted the speech patterns on the frequency scale, thus producing sounds which could not be perceived as speech while yet preserving quite exactly the acoustic variations that had, in the speech stimuli, cued the perceived linguistic distinction. In the discrimination of the control stimuli no peak in discrimination was found, either in the region corresponding to the location of the phoneme boundary or, indeed, in any other part of the stimulus continuum. This would indicate that the discrimination peak found with the speech stimuli is to be attributed to learning. It was apparent, further, that the discriminability of the non-speech controls was, at all points, below that of the speech. From this one would conclude that the learning effect consisted entirely of acquired distinctiveness.

The inverted patterns were not a perfect control. Nor was it possible that they could have been, since the ideal condition would have required that the controls be identical with the speech stimuli and yet not be perceived as speech. The control stimuli that were used in the experiment on /d,t/ had the salient shortcoming that the frequency of the formant whose time of onset varied was below the other two formants in the speech stimuli and above them in the controls. Since masking effects are greater from low frequencies to high, it is possible that the variations in onset were to some extent masked out in the control.

The specific purpose of the present experiment is to obtain data from an experiment analogous to the one just described, but with a more appropriate control. The linguistic distinction to be investigated is that between /b/ and /p/ in intervocalic position, specifically in the words *rabid* and *rapid*. In studying the perception of this distinction Lisker (1957; in preparation) has found that a sufficient cue—not the most important one, perhaps, but one that is nonetheless adequate—is the duration of the silent interval between the first and second syllables. When that interval is relatively short the listener hears *rabid*; increasing the interval causes the perception to change to *rapid*. The discriminability of such patterns, differing only in duration of the intervocalic silent interval, can, of course, be measured and then compared with the discriminability of the same durations of silence enclosed between two bursts of noise. These latter stimuli are particularly appropriate as non-speech controls, since they can be identical with the speech sounds, not only in the values of the stimulus variables (i.e., the duration of the silent interval), but also in regard to certain important constant aspects of the stimuli, such as the durations and amplitude envelopes of the sounds that bound the silent intervals. Comparing the discriminability of these speech and non-speech stimuli should help greatly to determine whether the

discrimination of the speech sounds reflects the effects of learning, and, if so, to discover which direction the learning has taken.

Procedure

One set of stimuli was generated from a hand-painted spectrogram like that shown in Fig. 1. When converted to sound by the Pattern Playback, this spectrogram produces a reasonable approximation of the word *rabid*. From Lisker's research (in preparation), we know that if the interval of silence between first and second syllables is made longer than that shown, the listener will hear *rapid*. To vary the duration of the silent interval and thus produce a series of sounds which would be perceived as *rabid* at one end and as *rapid* at the other, we made numerous magnetic tape recordings of the sound produced from the spectrogram shown in the figure, cut the magnetic tape in each case so as to separate the two syllables, and then inserted appropriate lengths of blank tape. In this way we created a series of 12 stimuli in which the silent interval varied between 20 and 130 msec. in steps of 10 msec. For convenience, we will refer to this set of sounds as the " speech stimuli " and designate each member of the set by the duration of the silent interval. Thus, Speech Stimulus 40 is that pattern which has 40 msec. of silent interval between the first and second syllables.

Our own listening convinced us that these particular stimuli did, indeed, sound like *rabid* or *rapid,* and that the shift from the one to the other occurred in the vicinity of 70 msec. Not unexpectedly, it appeared further that the longest silent intervals produced stimuli that would surely sound odd and unrealistic to speakers of English. We nevertheless included these extreme durations because we wanted to make certain that complete psychophysical functions would be obtained with the possibly less discriminable control stimuli to be described below.

It should be noted here that it is possible to begin with a recording of *rapid* as spoken by a human being and then, by reducing the duration of the interval between syllables, to convert it to *rabid*. The conversion is not wholly convincing, however, because there are several other cues to voicing besides the duration of the silent interval, and these are not changed as the interval is lengthened or shortened. Synthetic speech has the advantage here that it is possible to neutralize all the cues except the duration of the silent interval, and thus to produce a more satisfactory set of stimuli.

It was indicated in the introduction that we wanted as control stimuli a set of sounds as similar as possible to the speech series, and yet not perceivable as speech. To obtain such controls we used the speech stimuli to modulate noise signals and thus produce patterns consisting of noise-silent interval-noise in which the durations and rates of turn-on and turn-off would be the same as in the speech stimuli. The equipment and procedure for producing the control stimuli were as follows:[3]

The original speech stimuli were modulated by a 10 kc. carrier in a balanced

[3] *We gratefully acknowledge our debt to Dr. Carl Brandauer for devising the method of generating the control stimuli.*

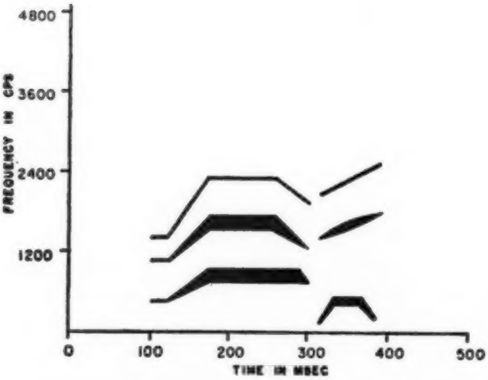


Fig. 1. Hand-painted spectrogram from which the stimuli of the experiment were produced.

modulator. The modulated signal was half-wave rectified, put through a low-pass filter (150 cps. cut-off), and this envelope waveform was then used to modulate a band-limited white noise (d.c. to 1500 cps.) in another balanced modulator. The circuit parameters were adjusted to give the best possible match to the envelopes of the original set of speech stimuli.

The extent to which we succeeded in matching speech and control stimuli was measured in two ways. First, we made a detailed visual comparison of the oscillograms of several pairs of stimuli. The envelopes of the speech and control stimuli were found to be very similar. As a second check, we made spectrograms on a Kay Sonograph of 36 pairs of speech and control stimuli and measured the duration of each silent interval. The averages for the two kinds of stimuli proved to be almost exactly the same. The variability of the control stimuli was somewhat greater, as we might expect, but the difference was not significant by an *F* test.

Subjects

All subjects in the experiment were undergraduate or graduate students at the University of Connecticut. They were paid volunteers with no special training in phonetics, and they were naive with respect to the purposes of the experiment.

There were 12 subjects in all, chosen from a group of 31 on the basis of a pre-test. The purpose of the selection was to insure that all subjects ultimately serving in the experiment would have a sharp and clear phoneme boundary.[4] In the pre-test, the

[4] *Since the experiment was intended to yield information on the relative discriminability of sounds within and across phoneme boundaries, only subjects with sharp boundaries were suitable.*

group of 31 subjects listened to the stimuli in various orders (approximately 28 presentations per stimulus) under instructions to identify each stimulus as *rabid* or *rapid*. On the basis of the data so obtained, we selected for service in the experiment the 12 subjects who had the sharpest phoneme boundaries. The results from this pre-test session were not otherwise used, and the data are not presented in the Results section. It should be noted, however, that all of the original group of 31 did reasonably well—so much so that the difference between the selected and rejected groups was very small.

Stimulus Presentation

As in the previous experiments in this series, the discriminability of both speech and control stimuli was measured by an ABX procedure—that is, the stimuli were presented in groups of three, and subjects were asked to determine, by whatever cues they could perceive, whether the third stimulus, X, was identical with the first stimulus, A, or the second stimulus, B. (In fact, X was always identical either with A or with B.) The measure of the discriminability of any pair of stimuli was, then, the proportion of the presentations on which the subject matched X correctly to A or B.

The A and B stimuli to be discriminated differed in silent interval by 20, 30, 40, 50, 60, 70, 80 and 90 msec. For example, Stimulus 20 was compared with Stimulus 40, 50, 60, 70, 80, 90, 100 and 110; Stimulus 30 was paired with Stimulus 50, 60, 70, 80, 90, 100, 110 and 120. The 10 stimulus comparisons in which pairs differ by 20 msec. will be called the 2-step series, the series of 9 pairs that differed by 30 msec. will be called the 3-step series, and so forth.

The total number of stimulus pairs is 52. Each stimulus comparison appeared in ABX triads of four forms—ABA, ABB, BAA, and BAB. For example, the four two-step comparisons for the 40-msec. stimulus were 40-60-40, 40-60-60, 60-40-40, and 60-40-60.

These triads were spliced together so as to form four test orders. Each stimulus comparison occurred once in each test order in one of its four forms. One second separated the members of a triad, while the triads were separated by four seconds. After the first four orders had been completed, a second set of four was made by shuffling each of the original orders. There were, then, eight test orders for the measurement of speech discrimination.

The non-speech stimuli were made into eight tapes in exactly the same fashion. That is, for each speech tape we made a control tape with a non-speech stimulus substituted for the speech stimulus having the same time separation between syllables.

The selected subjects listened to each of the eight speech and control tapes five times under discrimination instructions. Since a given stimulus comparison is presented once on each tape, the measure of discriminability at any point on each stimulus continuum is based on 40 determinations for each subject.

The purposes of the experiment required a comparison of the speech stimuli within and across the *rabid-rapid* phoneme boundary. Accordingly, we needed an accurate determination of the location of the boundary for the 12 subjects of the experiment.

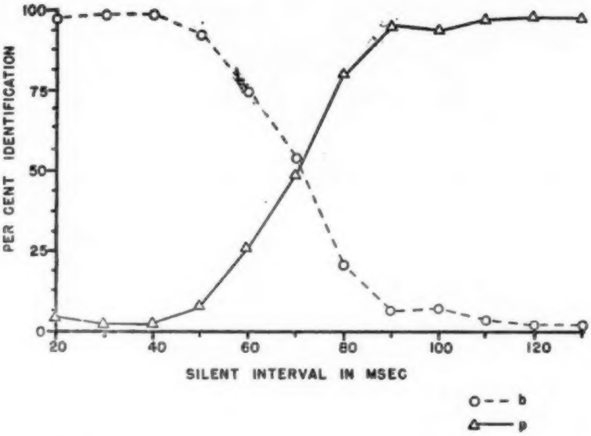


Fig. 2. Identification of the synthetic speech stimuli as /b/ or /p/, plotted against the duration of the silent interval between first and second syllables. The data are from the pooled responses of all 12 subjects.

To this end, the speech tapes were presented to each subject a total of 32 times with instructions to label each stimulus as *rapid* or *rabid*.

The whole experimental design, then, was set up so that each subject would perform three tasks: speech discrimination, noise discrimination, and phoneme labelling. A schedule was arranged for each subject such that the three tasks were distributed through all experimental sessions. Working in test sessions of about 20 minutes, each subject took about four months to complete the experiment.

Results

Phoneme Identification and the Location of the Boundary

Fig. 2 shows how the listeners assigned the phoneme labels /b/ or /p/ to the various stimuli. These functions which represent the pooled responses of all subjects, indicate that the phoneme identifications were made with fair consistency, and that the boundary between /b/ and /p/ lay at about 70 msec. of silent interval. It is also apparent from these data that the phoneme boundary is reasonably sharp.

Discrimination of the Speech Stimuli

The solid lines of Fig. 3 connect points which represent the percent correct discrimination of various pairs of speech stimuli at all values of the stimulus variable. As in the labelling functions of Fig. 2, the data from all subjects have been pooled.

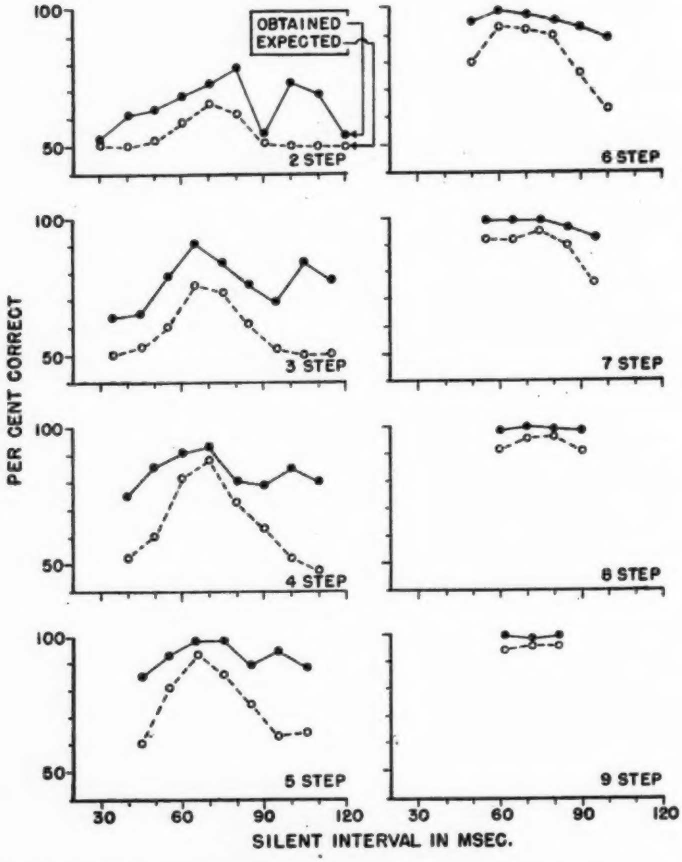


Fig. 3. Obtained and expected discrimination for the 1- through 9-step differences among the synthetic speech stimuli. The data are from the pooled responses of all 12 subjects.

For greater ease in reading the data, the graphs have been separated according to the amount of difference in silent interval by which the two stimuli are separated. Thus, in the first graph at the upper left, labelled " 2 step," the stimuli which were paired for discrimination (in ABX triads) always differed by 20 msec. (two steps on the stimulus scale) of silent interval. The first point on this graph indicates, then, that the subjects discriminated with 53% accuracy when the stimuli in the ABX triads had 20 and 40 msec. of silent interval. The next point shows that discrimination rose to 61% for the stimuli with 30 and 50 msec. of silent interval. Data points on

the graphs for the other stimulus comparisons, in which the differences between the stimuli ranged from three to nine steps, are to be read in similar fashion.

It is apparent, especially in the 2-, 3-, and 4-step graphs, that there are two peaks in the discrimination functions, a relatively large one near the centre of the stimulus continuum and a somewhat smaller one farther to the right. For the moment we will confine our attention to the larger peak.

Reference back to the phoneme identification data in Fig. 2 reminds us that the phoneme boundary is in the vicinity of 70 msec. and we see in Fig. 3 that the larger peak in the discrimination function occurs in this same region. This is to say, of course, that discrimination is better across the phoneme boundary than in the middle of the phoneme category. But instead of developing this comparison stimulus by stimulus, we will turn to a simple model developed in an earlier study (Liberman *et al.*, 1957) of this same problem, and evaluate the data with regard to the extent to which they fit that model.

Make the extreme assumption that the listeners can only hear these stimuli phonemically—that is, as /b/ or /p/—and can detect no other differences among them. Using the phoneme labelling data as a basis, one then predicts the accuracy with which the listener can be expected to discriminate all stimulus pairs. Thus, if the subject had always identified two stimuli as being members of the same phoneme class, he would be expected to discriminate the stimuli at a chance level. To the extent that he identifies two stimuli as belonging in different phoneme classes, he would, to precisely that extent, correctly discriminate them. In general, this assumption will predict peaks in the discrimination function wherever there are abrupt changes or inflections in the phoneme labelling curves, the height of the peak being a function of the abruptness and extent of the shift in phoneme labels. A more detailed description of the model and the derivation of the technique for predicting the discrimination functions are to be found in the earlier article (Liberman *et al.*, 1957).[5]

The discrimination functions that are predicted from the assumption of categorical perception are shown in Fig. 3 as the dashed lines. A comparison of these " expected " curves with the discrimination data that were actually obtained, and described earlier, leads to several conclusions. First, it will be noted that the left-hand portions of the two curves are reasonably similar in shape. This means that the variations in discriminability follow the change in phoneme labels. More specifically, it means that a discrimination peak does, indeed, occur at the phoneme boundary. The second conclusion is that the obtained functions tend in general to lie above the expected functions. To that extent the listeners are able in discriminating these stimuli to extract some information in addition to that which is revealed by the way in which they label the stimuli as phonemes.

[5] *In the earlier experiment (Liberman* et al., *1957) the labelling data were obtained by presenting the stimuli one at a time; in the present experiment the stimuli were presented for labelling in triads (see Procedure). Although the labelling functions were obtained by these different procedures, the calculation of the predicted discrimination values were made from the labelling data in exactly the same way in the two studies.*

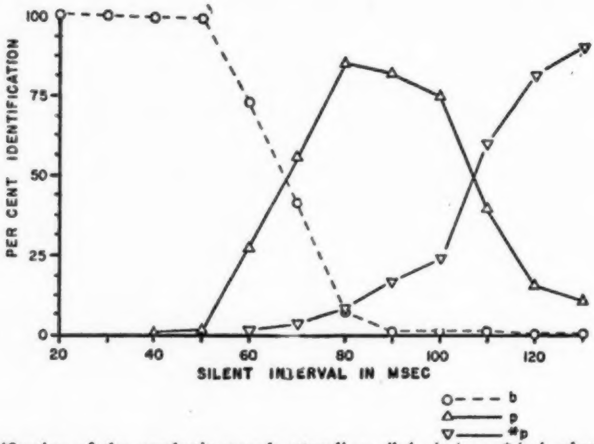


Fig. 4. Identification of the synthetic speech stimuli as /b/, /p/, or */p/, plotted against the duration of the silent interval. The data were obtained from the pooled responses of the seven subjects who served in this part of the experiment.

The Second Peak and a Third Category

Attention has been called to the fact that the obtained discrimination functions show a second peak at values of the stimulus variable greater than those at which boundary between /b/ and /p/ is located. On the assumption that this might imply a third category, we listened again to the stimuli and found that we had, indeed, carried the values of the silent interval to such extreme lengths as to have created, perhaps, an additional class of sounds.[6] This was a strange and unnatural /p/ to American ears, but we thought that it might nevertheless be heard and articulated by our listeners almost as if it were a different speech entity. We therefore recalled as many of the subjects as were still available (the number was seven), discussed the stimuli with them, and discovered that they, too, thought that some of them belonged in a separate class. To obtain more information about this third category we presented the stimuli (in random order, as before) to these subjects with instructions to identify each one as /b/, /p/, or */p/, the last named being the designation we chose for what we, and our subjects, had heard as the unnatural /p/. That the third category, */p/, did exist for these subjects is indicated by the graphs of Fig. 4. A comparison of these graphs with those that describe the results of the two-category judgments (Fig. 2) indicates that the distinction between */p/ and /p/ is not so clear as that between /b/ and /p/ in the original judgments (for which the subjects were allowed only the /b/ and /p/ categories). This is to be inferred from the fact that the curves

[6] *As was pointed out under Procedure, we were aware when we produced these speech sounds that curious effects could be heard at durations of silent interval greater than 100 msec., but we had decided to include these extreme stimulus values in order to be certain of obtaining complete psychophysical functions with the control stimuli.*

representing the /p/ and */p/ judgments (in Fig. 4) do not rise to 100% as the /b/ curve does, and as both the /b/ and /p/ curves do in the two-category situation (Fig. 2). Nevertheless, the labelling data show that a */p/ category does exist, and they provide a basis for predicting a new set of discrimination results from the assumption of categorical perception. These predictions are shown as the dotted lines in Fig. 5, together with the discrimination data that were actually obtained with the seven subjects who made the three-category judgments. There is a second peak in the expected discrimination function corresponding to the boundary between the second and third categories of Fig. 4. Moreover, this second peak fits moderately well the second peak in the obtained discrimination functions. Clearly, the expected and obtained functions now agree somewhat better than before, but there remains a constant difference between the functions in the direction of better discrimination than is to be expected on the extreme assumption of categorical perception. We will not try to specify the magnitude of the discrepancy in terms of some single meaningful quantity, because we are not prepared at this stage to decide which of several possible measures is best. We will only say that the discrepancy is somewhat greater here than it was in the studies of /b,d,g/, /d,t/, and /sl,spl/, where the perception was more nearly categorical; also, that it is less than in the vowels and prosodic features, where perception was essentially non-categorical, i.e., continuously changing with progressive changes in the stimulus.

In terms of the theory outlined in the introduction, perception of speech becomes linked to the feedback from the articulatory movements the listener makes in speaking. We should expect, then, that perception would be completely categorical (i.e., that the discrimination functions would show a peak at the phoneme boundary and be perfectly predictable from the phoneme labelling data) if the listener makes exactly the same articulatory response to the various stimuli to which he attaches the same phoneme label, and very different articulatory responses to sounds he calls by different phoneme names. At the other extreme, speech perception would be expected to be perfectly continuous (i.e., the discrimination function would show no peak at the phoneme boundary and might lie at a level far higher than that which is predicted from the phoneme labelling data) if the listener's mimicking articulations change in linear fashion with variations in the acoustic stimuli, both within and across phoneme boundaries.

One might expect something like the results we found in this experiment—perception which is almost categorical, but not quite—if it be the case that the articulatory response changes most rapidly at the phoneme boundary, but that there is, nevertheless, some small variation within the phoneme class. In an attempt to find out whether or not this was so, we had several of the listeners try to mimic the various stimuli, and then undertook to measure the duration of the silent interval between the two syllables. It proved to be difficult to obtain highly reliable measurements, chiefly because the subjects produced variations in other acoustic features, such as the first-formant transition and presence or absence of voicing, which are more important than

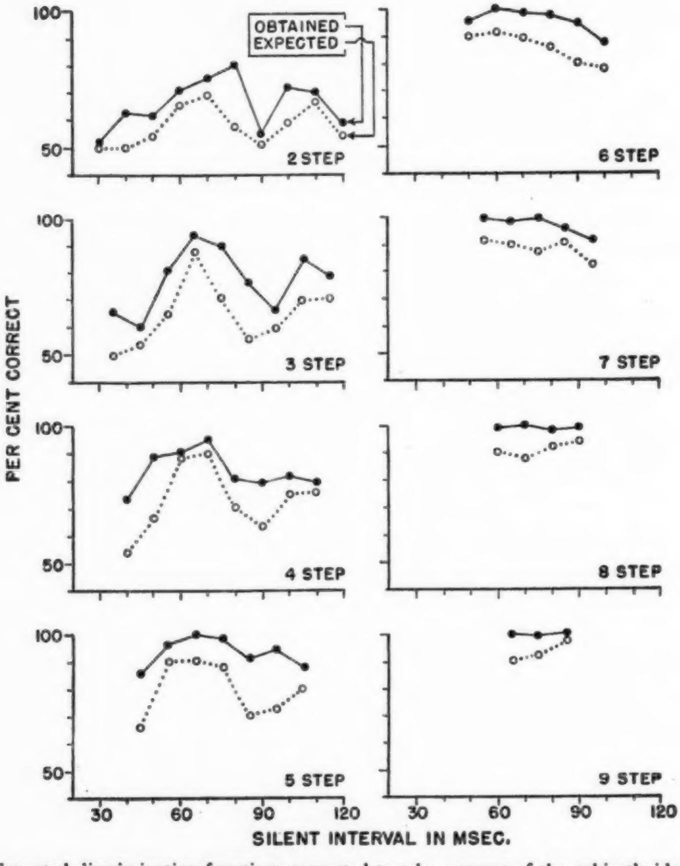


Fig. 5. Expected discrimination functions corrected to take account of the subject's identification of the third category, */p/, together with the obtained discrimination functions. The data are from the pooled responses of the same seven subjects whose three-category identification functions are shown in Fig. 4.

duration of silent interval as such, and which tended to obscure it. When, in the course of this work, it became apparent that we would be able to get far more precise mimicry data in the study of other phonemic distinctions,[7] we abandoned the attempt to measure the mimicry of *rabid-rapid*.

[7] *One such study has been completed since this paper was written and has been published as an abstract, see Harris, Bastian and Liberman (1961). It is now being prepared for regular publication.*

In some respects, then, this experiment yielded less than we might have wished. We should remember, however, that we undertook it because, being interested in the discrimination peaks which sometimes occur at phoneme boundaries, we wanted a fair comparison between the discrimination of an acoustic variable when it cues a phonemic contrast and when, in a non-speech context, it does not. Fortunately for that purpose, the discrimination of the speech stimuli does have a peak (or actually two) sufficiently high to make the comparison with the non-speech control an interesting one.

Discrimination of Noise Control

In the stimuli used as controls, bursts of noise served to bound silent intervals that duplicated those of the speech stimuli. It is also relevant to recall that the noise bursts were matched with the speech syllables in regard to such constant features as amplitude envelope and duration.

The discrimination results obtained with the noise control are shown in Fig. 6. For comparison the results obtained with the speech stimuli and previously shown in Fig. 3 are also presented.

One sees immediately that the discrimination peaks of the speech stimuli are much higher and sharper than any peaks which appear in the control data. More generally it is clear that the discriminability of the speech sounds is considerably greater than the control at most points. At a few values of the stimulus variable the two are equal, or very nearly so. Out of a total of 52 points at which the two sets of curves can be compared there is only one at which the speech discrimination dips below the control. In this one case the difference is small, and probably not at all reliable. If the noise stimuli are truly an appropriate control—that is, if they fairly represent the discriminability of the speech stimuli prior to linguistic training—we may conclude that the results obtained with the speech stimuli reflect the effects of a very considerable amount of learning. It is clear, further, that the entire learning effect consists of a sharpening of discrimination in the vicinity of the phoneme boundary. There is no indication that discrimination of the speech has been reduced (below the control) within the phoneme category. In terms of the psychological processes discussed in the introduction, we should say that there is here a very considerable amount of acquired distinctiveness, but no acquired similarity.

Reference has been made to the earlier study of /d,t/, in which discrimination of variations in a cue for a phonemic distinction were compared with discrimination of the same variable in a non-speech context. It was pointed out that the acoustic variations might have been masked in the control stimuli, and the conclusion drawn from a comparison of speech and control discrimination was, therefore, thought to be open to question. We should note that the results of the present experiment, with its more adequate controls, agree with those of the earlier study in that there was evidence of a large amount of acquired distinctiveness and no acquired similarity.

We ought, perhaps, to remark on the fact that this experiment and the earlier one on /d,t/ differ somewhat in regard to the finding of no reduction in discrimination within the phoneme category (acquired similarity). In the earlier experiment the

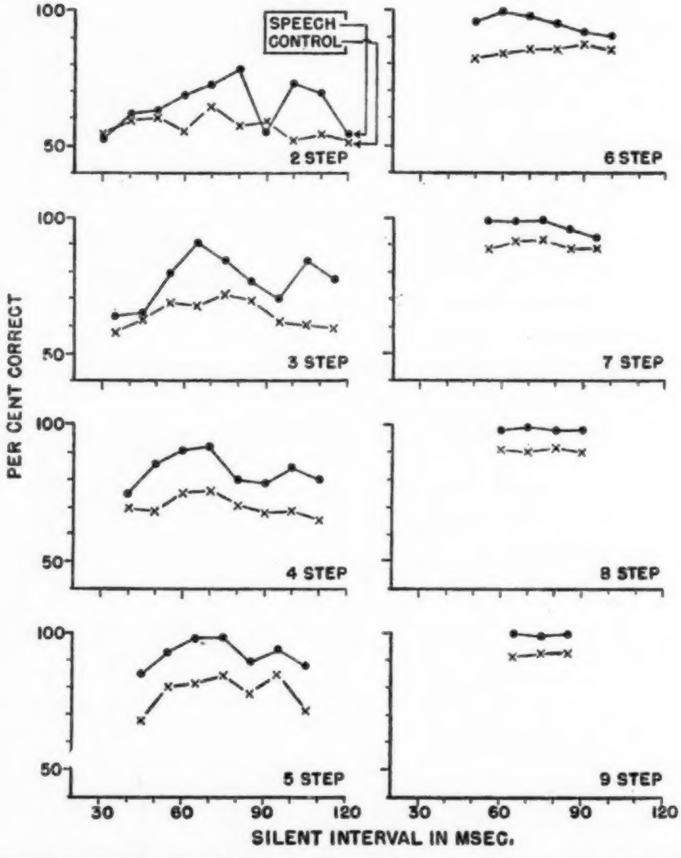


Fig. 6. Discrimination functions for the 1- through 9- step differences among the non-speech control stimuli. The obtained speech discrimination functions previously shown in Fig. 3 are reproduced here to facilitate comparison. For both sets of functions, the data are from the pooled responses of all 12 subjects.

discriminability of the non-speech stimuli was very poor, lying generally at or just slightly above chance. There was, then, no room for a process like acquired similarity to show itself, for no amount of training could possibly have reduced discrimination much further. In the present experiment the discrimination of the noise control stimuli rose to higher levels. To determine just how much room this provided for acquired similarity, we must compare the three discrimination functions previously presented: the obtained discrimination of the speech sounds, the expected discrimination of the speech sounds, and the obtained discrimination of the noise control. For that purpose

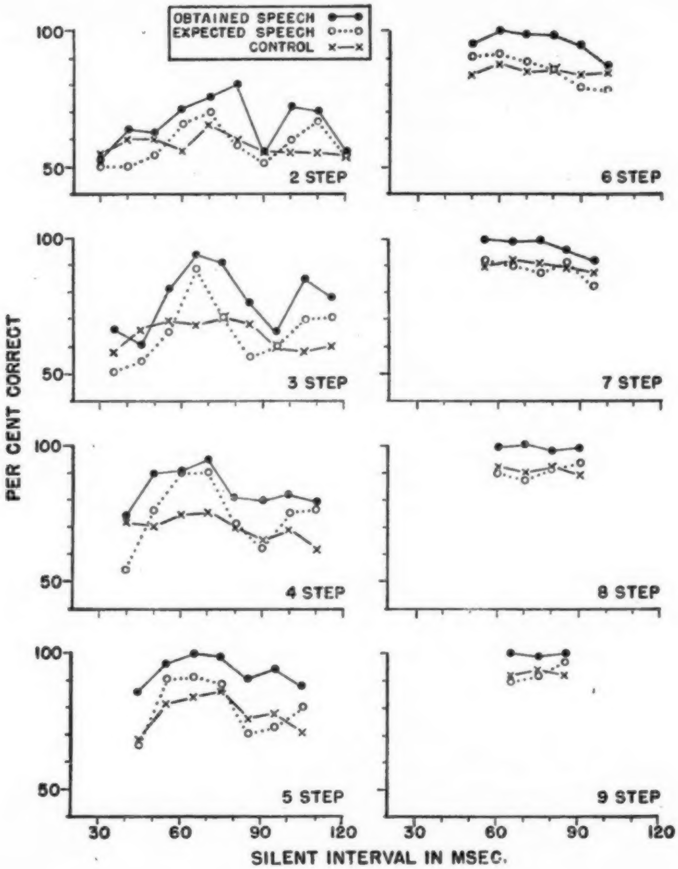


Fig. 7. Obtained and expected speech discrimination functions, together with the discrimination functions for the non-speech control stimuli. The obtained and expected speech functions are the same as those shown in Fig. 5. All data are from the pooled responses of the seven subjects who provided the results shown in that figure.

the three functions are shown together in Fig. 7. (We here use the data for the seven subjects who made the three-category judgments, because these data most adequately depict the relationship between discrimination and phoneme labelling.)

Within the /b/ category—that is, on the left-hand side of the graphs—we see in the 2-, 3-, and 4-step data that the discriminability of the noise does lie somewhat above the predicted discriminability of the speech. This means that the original discriminability of speech may be presumed to have been greater than it needed to be

to meet the requirements of the linguistic situation. It also means that if the listener makes the same articulatory response to these stimuli, we should expect, according to theory, that the discrimination of the speech would have been reduced below the noise control, down to the predicted level. We see that this has not happened, and we conclude that we have here a case in which acquired similarity did not occur, though there was room for it. Beyond the four-step comparisons practically all stimulus pairs fall across a phoneme boundary; discrimination is therefore predicted in general to be at fairly high levels, well above the noise control at all points, and we cannot make the kind of test we are here considering. Between the second and third categories (/p/ and */p/) the noise discrimination again lies above the expected values, and we find, as we did in the similar situation within the /b/ category, that the obtained speech discrimination has not been reduced to the expected level.

While the relevant data are not very clear or compelling, there is a certain amount of evidence that acquired similarity might have occurred but did not. Whether it *should* have occurred, according to theory, is a separate question, and one that is not readily answered because the critical mimicry data are missing. To see why these data are critical, let us imagine that reliable mimicry measurements have been obtained, and then consider the implications of different kinds of results. Suppose first, that the subjects are found to make the same articulatory response in mimicking the sounds within the phoneme category; the theory as it now stands demands that they be unable to discriminate these sounds. Given this outcome of the mimicry experiment, and given the fact that acquired similarity did not occur though there was room for it, we should have to modify the theory. It would appear then that while the articulatory responses become involved in the perception of speech, the listener still has some choice remaining to him: if falling back on the feedback from articulation serves to sharpen discrimination (as it apparently does at the phoneme boundary), the listener takes advantage of this possibility and discriminates better than he would otherwise have done; if the articulatory feedback has the effect of dulling discrimination, the listener effectively ignores it, responds directly to the acoustic stimulus, and suffers no loss in acuity. This would say, in general, that the acquired similarity paradigm describes an event which does not occur, at least not in speech perception; we should assume then that while we may, on occasion, find it necessary to disregard clearly perceptible differences, and for very practical reasons to call distinguishable stimuli by the same phoneme name, we do not as a consequence really lose our ability to discriminate.

The alternative result of the mimicry experiment would be to show that the listener can and does mimic some of the stimulus changes within the phoneme category—that while the articulatory response changes most rapidly at the phoneme boundary, there is nevertheless some variation in mimicking sounds the subject calls by the same phoneme name. In that case, we should not expect discrimination within the phoneme class to be reduced to chance, and, depending on the particular nature of the mimicry results, the theory in its present form might be rather precisely confirmed. We hope

that more light will be shed on this question in research on other phoneme distinctions where mimicry can be more easily measured.

It will be recalled that the noise control stimuli were produced in such a way as to make their amplitude envelopes and durations approximate the speech signals as closely as possible. This is surely the most appropriate way to obtain control signals for use with the speech stimuli of this experiment, but it may raise a question about the extent to which the results represent what would be obtained with the simpler and more regular stimuli that are usual in psychophysical studies. To answer this question, we prepared a new set of control signals which had intervals of silence like those of the original controls, bounded by segments of noise which differed from the original in having abrupt onsets and offsets (produced by cutting the magnetic tape at a 90-degree angle) and in being of equal duration (300 msec.). On testing the discriminability of these stimuli with the same subjects who had served in the earlier parts of the experiment, we obtained results very similar to those found with the original set of controls.

As a further test of generality we undertook to obtain some indication of the effect, if any, of the particular psychophysical procedure (ABX) that had been used. For that purpose we measured the discriminability of the new noise stimuli by the ABX method and also by an adaptation of the forced-choice temporal interval method developed by Blackwell (1952) for measuring visual thresholds. In the latter procedure, as in ABX, the stimuli are presented in triads composed of two stimuli which are identical and one which is different, but the "different" stimulus can appear in any of the three positions (first, second, or third) of the triad, and the subject's task is to tell where, i.e., in which position, it is. When the data obtained by the two methods were adjusted to take account of the different levels of chance performance (50% in ABX and 33⅓% in the new method), the levels of discrimination proved to be the same.

We feel reasonably certain that the original noise stimuli are appropriate controls. The two pulses of noise were carefully matched in envelope and duration with the two syllables of the speech stimuli, and the interval of silence, which was in both speech and control the variable part of the pattern, is out in the open, as it were, where it is not likely to be masked or otherwise interfered with. Moreover, as we have seen, the data obtained with these matched controls would appear to have some generality, since very similar results were found with more standard patterns and, indeed, with a different psychophysical method. We will assume, then, that the discrimination functions obtained with the noise controls approximate the discrimination of the speech patterns as they would have been without linguistic experience. The fact that the peaks of the speech discrimination functions rose above the control may then be taken to indicate a learned increase in discrimination across phoneme boundaries. The speech functions nowhere fell below the control, from which we conclude that there was, for whatever reason, no loss in discrimination within phoneme categories.

REFERENCES

ABRAMSON, A. S. (1961). Identification and discrimination of phonemic tones. *J. acoust. Soc. Amer.*, **33**, 842 (Abstract).

ABRAMSON, A. S. and BASTIAN, J. (in preparation). Identification and discrimination of phonemic vowel duration.

BASTIAN, J., DELATTRE, P. and LIBERMAN, A. M. (1959). Silent interval as a cue for the distinction between stops and semivowels in medial position. *J. acoust. Soc. Amer.*, **31**, 1568 (Abstract).

BASTIAN, J., EIMAS, P. D. and LIBERMAN, A. M. (1961). Identification and discrimination of a phonemic contrast induced by silent interval. *J. acoust. Soc. Amer.*, **33**, 842 (Abstract).

BLACKWELL, H. R. (1952). Studies of psychophysical methods for measuring visual thresholds. *J. opt. Soc. Amer.*, **42**, 606.

CHAPANIS, A. S. and HALSEY, R. M. (1956). Absolute judgments of spectrum colours. *J. Psychol.*, **42**, 99.

COOPER, F. S., DELATTRE, P. C., LIBERMAN, A. M., BORST, J. M. and GERSTMAN, L. J. (1952). Some experiments on the perception of synthetic speech sounds. *J. acoust. Soc. Amer.*, **24**, 597.

COOPER, F. S., LIBERMAN, A. M., HARRIS, K. S. and GRUBB, P. M. (1961). Some input-output relations observed in experiments on the perception of speech. *Proceedings of the Second International Congress on Cybernetics*. Namur, Belgium.

DELATTRE, P. C., LIBERMAN, A. M. and COOPER, F. S. (1955). Acoustic loci and transitional cues for consonants. *J. acoust. Soc. Amer.*, **27**, 769.

FRY, D. B., ABRAMSON, A. S., EIMAS, P. D. and LIBERMAN, A. M. (in preparation). The identification and discrimination of synthetic vowels.

GARNER, W. R. (1953). An informational analysis of absolute judgments of loudness. *J. exp. Psychol.*, **46**, 373.

GRIFFITH, B. C. (1958). A study of the relation between phoneme labelling and discriminability in the perception of synthetic stop consonants. *Unpublished Ph.D. dissertation, University of Connecticut.*

HARRIS, K. S., HOFFMAN, H. S., LIBERMAN, A. M., DELATTRE, P. C. and COOPER, F. S. (1958). Effect of third-formant transitions on the perception of the voiced stop consonants. *J. acoust. Soc. Amer.*, **30**, 122.

HARRIS, K. S., BASTIAN, J. and LIBERMAN, A. M. (1961). Mimicry and the perception of a phonemic contrast induced by silent interval: electromyographic and acoustic measures. *J. acoust. Soc. Amer.*, **33**, 842 (Abstract).

LADEFOGED, P. (1959). The perception of speech. In *Mechanisation of Thought Processes*, Vol. I (London).

LIBERMAN, A. M., DELATTRE, P. C. and COOPER, F. S. (1952). The role of selected stimulus variables in the perception of the unvoiced stop consonants. *Amer. J. Psychol.*, **65**, 497.

LIBERMAN, A. M., DELATTRE, P. C., COOPER, F. S. and GERSTMAN, L. J. (1954). The role of consonant-vowel transitions in the perception of stop and nasal consonants. *Psychol. Monogr.*, **68**, No. 8 (Whole No. 379).

LIBERMAN, A. M., HARRIS, K. S., HOFFMAN, H. S. and GRIFFITH, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *J. exp. Psychol.*, **54**, 358.

LIBERMAN, A. M. (1957). Some results of research on speech perception. *J. acoust. Soc. Amer.*, **29**, 117.

LIBERMAN, A. M., DELATTRE, P. C. and COOPER, F. S. (1958). Some cues for the distinction between voiced and voiceless stops in initial position. *Language and Speech*, **1**, 153.

LIBERMAN, A. M., HARRIS, K. S., KINNEY, J. A. and LANE, H. (1961). The discrimination of relative onset-time of the components of certain speech and non-speech patterns. *J. exp. Psychol.*, 61, 379.

LISKER, L. (1957). Closure duration and the inter-vocalic voiced-voiceless distinction in English. *Language*, 33, 42.

LISKER, L. (in preparation). On separating some acoustic cues to the voicing of intervocalic stops in English.

MILLER, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychol. Rev.*, 63, 81.

POLLACK, I. (1952). The information of elementary auditory displays. *J. acoust. Soc. Amer.*, 24, 745.

POLLACK, I. (1953). The information of elementary auditory displays. II. *J. acoust. Soc. Amer.*, 25, 765.

RELATIONSHIPS AMONG FUNDAMENTAL FREQUENCY, VOCAL SOUND PRESSURE, AND RATE OF SPEAKING*

JOHN W. BLACK
Ohio State University

Twenty males who could control their vocal effort to reach specified soft and loud vocal levels, spanning 30 db., practised and recorded three vowels and three phrases at four levels, ranging from soft to loud.

Increments in vocal effort were accompanied by increase in fundamental frequency, the latter shifting upward increasingly with successive steps in sound pressure.

The vocal changes that occurred from one level of speaking to another were somewhat specific to the material that was spoken.

Phrases that were spoken with different amounts of vocal effort, soft to loud, were spoken at the slowest rate when said softly.

Both *vocal effort* and *sound pressure level of speech* are used to describe the same aspect of talking. The former is obviously subjective and has physiological connotations; the latter is implicitly a physical dimension and exact. Actually, the variability of the power of speech from moment to moment is so great that a " degree " of *vocal effort* gives almost as precise a description as an " amount " of *sound pressure level.*

As vocal effort is increased, the fundamental frequency of voice typically rises[1] and the rate of talking may be altered.[2] The present study was planned to examine these relationships and possibly to quantify them.

Six samples of speech were used: three vowels [ɑ,o,i] and three short phrases (*take off runway five . . . stop at Corry Field . . . pull the setting back*).

Twenty young adult males served individually as experimental subjects. A subject, sitting comfortably, maintained a constant position relative to two microphones, one 18 in. removed from the speaker's mouth and the other at the corner of his mouth, out of the breath stream. The former microphone fed a General Radio sound level meter and the latter, through an attenuating network, an Ampex 400 recorder. The speaker practised saying the vowels and phrases at four degrees of vocal effort, ordered from

* *This work was conducted as a part of a contract between the Office of Naval Research and the Ohio State University Research Foundation.*

The experiment was conducted two times. Experimental error was suspected in the instance of the first data because of apparently excessive variance among the measures. The second set of data confirmed the earlier set.

[1] *Black, J. W. and Morrill, S. N. (1954).* The pitch of sidetone. *Pensacola: Joint Project, The Ohio State University Research Foundation and U.S. Naval School of Aviation Medicine. Joint Project Report No. 31. The Bureau of Medicine and Surgery Project No. 001 064 01 31.*

[2] *Lightfoot, C. (1949).* Effects of the mode and levels of transmitting messages upon the relationship between their duration and the duration of their repetition. *Kenyon College, Technical Report—SDC 411-1-5.*

soft to loud. His principal instruction was to try to register 70, 80, 90, and 100 db. on the sound level meter, C-scale (18 in. distant). The levels were selected after trying out a number of speakers to find feasible limits and after trying alternative instructions that were worded in terms of degrees of vocal effort. The order between vowels and phrases and among the samples of each was varied from speaker to speaker.

The experimenter monitored the level indicator and adjusted the relative gain of the recorder from condition to condition to +10, 0, −10, and −20 db. in order to compensate for the different levels of input.

The recordings were copied to a Bruel and Kjaer power level recorder, 10 mm./sec. These tracings yielded three measures: (a) the duration of each phrase, (b) an index of the relative sound pressure level of the phrases, and (c) similar measures of the level of the initial second of each vowel. A further copy of the recorded speech, slowed down by a factor of eight and made on an Edin pen-writing oscillograph, provided a wave-by-wave resolution of the speech suitable for determining the average fundamental frequencies of each vowel and phrase. Thus, the five sets of values shown in Tables 1 and 2 were obtained: (a) relative sound pressure of the three vowels, (b) mean fundamental frequency of each of the three vowels, (c) relative sound pressure of the phrases (determined as the mean of the three maximum deflections of the Bruel-Kjaer tracing of each recorded phrase), (d) duration of the phrases, and (e) mean fundamental frequency of the first 0·2 sec. of vocalization of each phrase. These five sets of measures were treated statistically in a series of five similar analyses of variance. Each was a triple analysis: (samples × vocal levels × subjects).

Tables 1 and 2 summarize the five analyses and enumerate the mean values associated with the four levels and the different samples. The gain of the recorder having been adjusted to compensate for changes in vocal level from one condition to another, small variance was anticipated in connection with vocal effort. This was the outcome with sustained vowels, sounds that are relatively constant from moment to moment; it was not the case with phrases. However, the *samples* × *levels* (AC) interaction values of Table 1 are noteworthy, particularly in the instances of the vowels. Obviously the three vowels did not maintain a constant relationship to one another in sound pressure at four levels of vocal effort. This fact emphasizes the singularity of various speech samples. The interaction is again apparent in the entries of Table 2. Two of the vowels [ɑ,o] are approximately 32 db. higher in level in Condition 4 than in Condition 1, while [i] varies only 27 db. from the condition of least vocal effort to the one of greatest effort.

The increments in fundamental frequency that accompanied the changes in level were not the same for the three vowels: [ɑ] increased 12·1 semitones from the condition of least effort to the one of greatest effort; [o], 12·6 semitones; and [i], 14·1 semitones. This between-sample variability in fundamental frequency was not evident among the phrases; in fact, the small variance associated with the fundamental frequency of the phrases led to pooling the three measures at each level, evident in Table 2.

Table 1

Source of Variance		*df*	VOWELS: *Relative Sound Pressure Level*	VOWELS: *Fundamental Frequency*	PHRASES: *Relative Sound Pressure Level*	PHRASES: *Fundamental Frequency*	PHRASES: *Duration*
Samples (vowels or phrases):	A	2	643.50*	39.96*	106.00*	0.15	239.16*
Subjects:	B	19	61.58	35.85	151.26	53.96	33.38
Levels:	C	3	9.33	2897.56*	106.33*	2894.42*	20.31*
AB		38	11.32	2.33	6.18	1.06	2.54
AC		6	18.83*	12.21*	7.50*	0.38	2.91*
BC		57	7.25	12.53	15.12	13.71	3.09
ABC		114	3.08	1.97	2.84	0.69	0.84

* Statistically significant beyond the 5% level of confidence.

Mean-square values of five analyses of variance, vowels and phrases, spoken under four sound pressure levels.

Table 2

		Relative Sound Pressure Level of Vowels (db.)			*Fundamental Frequency of Vowels* (cps.)			*Relative Sound Pressure Level of Phrases* (db.)			*Fundamental Frequency of Phrases (pooled)* (cps.)	*Duration of Phrases* (sec.)		
		[ɑ]	[o]	[i]	[ɑ]	[o]	[i]	1	2	3		1	2	3
Soft:	0 db.	0.9	0.9	0	109.7	109.7	110.2	0.9	0.1	0	111.8	1.32	1.33	1.12
Plus:	10 db.	14.6	11.8	8.2	121.8	123.0	125.2	10.0	8.2	7.9	129.2	1.26	1.22	1.00
Plus:	20 db.	23.7	21.1	17.5	156.2	160.6	166.2	19.2	17.3	16.7	167.4	1.29	1.21	1.02
Plus:	30 db.	32.9	33.2	27.4	222.7	233.5	250.9	29.0	27.6	25.5	237.1	1.31	1.20	0.99

Mean values of five sets of measures related to four levels of talking.

Soft voice or minimum vocal effort was accompanied by a slower rate of reading the phrases than was used with the other three levels of vocal effort. The mean durations of the phrases in the four incremental conditions of level were 1·26, 1·16, 1·17, and 1·17 sec.

The principal observations that emerge from this study follow. First, approximately equal increments in vocal sound pressure are accompanied by successively greater increments in fundamental frequency and within a 30 db. range the voice rises in excess of an octave in pitch. Since the instructions to the speaker were in terms of sound pressure level, not vocal effort, the increments in effort were probably unequal; this inequality was possibly reflected in the increments in fundamental frequency from one level to another. This was 13 cps. from the first to the second level; 38 cps. from the second to the third; and 74 cps. from the third to the fourth. These values are taken from the descriptive data for vowels in Table 2, but they are essentially the same for the phrases. Second, as vowels are spoken at each of a succession of sound pressure levels the accompanying fundamental frequencies are somewhat unique from one vowel to another. This can be considered as an interaction between the physiological effort of saying a particular vowel and the further effort that accompanies gradations of intensity of speech. Third, although the physiological limits for soft and loud speech were not reached by all speakers, these limits were approached for the group as a whole. The fundamental frequencies of Level 2 approximate ones that are frequently reported for adult male voices. One can probably assume that the level of speech that is usually studied as a normal or natural level is approximately 10 db. above minimum vocal effort for speech and about 20 db. below maximum vocal effort. Fourth, a slow rate characterizes soft speech.

AUTOMATIC SPEECH RECOGNITION PROCEDURES*

GORDON E. PETERSON
University of Michigan

This paper is concerned with the transformation of the varying acoustical parameters of speech to a discrete code to form the printed output of an automatic speech recognizer. The development of general automatic speech recognition procedures requires a definition of the linguistic code to be transcribed, and a statement of the dialectal and other conditions under which the recognition is to be achieved.

Essential procedures in automatic speech recognition include: the analysis of the input speech wave into a series of basic acoustical parameters in frequency; the representation of the normalized parameters by a set of phoneme and prosodeme candidates by reference to stored linguistic information; and the print-out into words separated by spaces and grouped by means of a set of punctuation marks. The possibility is considered of employing values of conventional spelling and punctuation in the automatic representation of spoken American English.

INTRODUCTION

Automatic speech recognition is primarily concerned with the conversion of the continuous functions of speech to a discrete symbolic representation. It is the objective of the present paper to discuss the general procedures and problems involved in making such a conversion. For basic concepts, the present paper relies heavily upon a previous development by Peterson and Harary (1961), and several terms are employed here in the manner defined in that reference.

According to the above indicated paper, phonetically similar phones are phones which have the same vowel or consonant parameter values. Two phones are functionally similar if they have phonetically similar environments and occur in semantically equivalent utterances. A phonetically related set is a set of phones obtained by taking the union of a maximal set of phonetically similar phones and all sets of functionally similar phones which contain a member of that set. The various languages of the world involve a large number of different phonetically related sets; any given language, however, involves only a relatively limited number of such phonetically related sets.

* *This research was supported by the United States Air Force Office of Scientific Research under Contract AF 49 (638)-492. The author would like to express his appreciation to Miss June E. Shoup and Dr. W. S-Y. Wang for their many helpful suggestions. This is a revision of a manuscript previously published under the title* "Linguistic Concepts in Automatic Speech Recognition Procedures" *in the* Proceedings of the Seminar on Speech Compression and Processing (Vol. 2), *sponsored by the Electronics Research Directorate of the Air Force Cambridge Research Center, Air Research and Development Command, 28-30 September, 1959.*

A phonetic recognizer which could distinguish among all possible phonetically related sets would be extremely complex, but even within a single language the number is far too large to make it practical to symbolize each different phonetically related set by a separate symbol at the output of an automatic speech recognizer. Rather, the objective is to approximate a phonemic transcription (as defined in the above mentioned paper).

Since it is very difficult to define the limits of a language, it appears impractical to consider the construction of a general speech recognizer for any input within a particular language. As a minimum, however, it would seem that an automatic speech recognizer should respond properly to various speakers of a particular dialect of a given language. Machines designed for more restricted inputs, where the redundancy is higher, may of course encompass a broader range of dialects, i.e. as vocabulary is decreased dialectal range may be increased.

In automatic speech recognition the speech wave provides the physical data which must be interpreted. Certainly, the speech wave does not contain the complete information necessary for its interpretation, for in generating and perceiving speech both the speaker and the listener have large amounts of stored information available to them. Thus there are two general types of information which are required to determine the output of an automatic speech recognizer: the input speech wave and stored information.

There are doubtless many different detailed procedures which may be employed in relating these two types of information to determine the output of a speech recognizer. At one extreme, the stored information might be employed to generate simulated speech waves, and these waves could then be compared with the waves to be identified. The simulated wave which had a minimum difference from the wave to be recognized would be accepted as a correct identification. It seems obvious, however, that for any one recognition it would be necessary to generate a large number of different waves, and that it would be difficult to determine when a minimal difference had been achieved.

In a more restricted approach by speech simulation, stored information might be employed to generate various speech elements to be compared with the signal at successive stages of its analysis. At an early stage of the recognition process, for example, input speech spectra could be compared with simulated spectra generated by the recognizer; simulated spectra having minimal differences from the input spectra would represent correct answers. Various types of simulations would be required at successive stages of the recognition process, and optimal identifications could be found by a series of successive approximations. When various detailed aspects of speech are generated for comparison with corresponding aspects of the input speech for the purpose of automatic speech recognition, the procedure may be called *analysis by speech simulations*.

In automatic speech recognition it is necessary to perform a sequence of identification operations, with each operation in the sequence based upon relevant stored information. The concept of stored information is employed here in a broad sense. In addition to stored representations of linguistic elements (as phonemes or words), such devices as

passive filter sets and switching networks represent stored information. Various types of stored information are required at successive stages of the recognition process: for example, speech spectra, phoneme sequences, words, etc. If the information is stored in circuit design and computer memory, it may be used for reference at successive stages of the recognition process. When various aspects of the speech signal are compared with stored elements and distributions in automatic speech recognition, the procedure is that of *dynamic speech analysis*.

It should be evident that the procedures of analysis by simulation and of dynamic speech analysis require essentially the same stored information and a similar sequence of operations for automatic speech recognition. They differ in instrumental detail, and possibly in recognition efficiency or accuracy. It is the approach of dynamic analysis which seems most reasonable to the present author and which will be developed in the present paper.

Acoustical data in automatic speech recognition

While the speech code is largely organized in terms of the capabilities and constraints of the physiological mechanism, it is the acoustical form of the speech signal which is primarily available for analysis. Under favourable conditions of speech transmission a high degree of intelligibility can be achieved with only the acoustical form of the speech signal, and this fact is ample evidence that the acoustical form of the signal is an adequate input for automatic speech recognition.

As is well known, acoustical speech waves involve a multi-dimensional continuum. The variations observed in acoustical speech waves reflect variations in the physiological production of speech (Fant, 1958), and so there is no basis for assuming a higher degree of quantization in physiological speech production than in the acoustical form of the speech signal. The transformation from the physiological production of speech to acoustical speech waves is highly complicated. Since several different physiological formations may map into the same acoustical time pattern, the correspondence is essentially many-to-one in nature. Physiological formations which map into the same acoustical time pattern form an acoustical equivalence class.

It follows from the above considerations that it is not the problem of automatic speech recognition to perform an inverse transform on the acoustical speech wave to reconstruct the physiological details of speech production. Rather, the essential problem is to map a series of discrete code elements on to the multi-dimensional continuum of the acoustical speech wave. Obviously, such a mapping will be more consistent if there is a clearly defined relationship between the properties of the acoustical speech wave and the elements of the discrete code with which it is represented.

A set of discrete elements, i.e. a code, may be used to provide a mathematical (and linguistic) representation of speech production. The transformation from speech

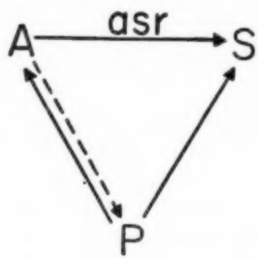


Fig. 1. Directed graph illustrating two different approaches to automatic speech recognition.

production (P) to discrete symbols (S) may be symbolized by P→ S, and the transformation from speech production (P) to speech acoustics (A) may be symbolized by P→ A. As illustrated in Fig. 1, it seems more reasonable to perform the transformation A→ S for purposes of automatic speech recognition, than to attempt to follow the more devious route of A→ P→ S. Thus the essential problem of automatic speech recognition is to perform logical operations on the acoustical speech parameters to achieve a direct representation of the phonemes and prosodemes of the acoustical speech signal by means of a discrete code. Since the properties of this code are determined by the physiological speech parameters, however, it is obvious that the development of a set of logical operations for interpreting the acoustical speech waves will be considerably aided by reference to the processes of speech production.

Speech waves. It is not the objective of automatic speech recognition to identify speech under all conditions of acoustics and conversational interchange. Where adverse acoustical conditions are imposed, it seems necessary to place a corresponding restriction upon the vocabulary to be recognized. The use of the basic procedures appropriate to general automatic speech recognition, however, should ensure greatest success in the case of limited vocabulary applications.

It also seems reasonable to expect the speaker to make some adjustments to the equipment. But much of the purpose of automatic speech recognition would be defeated if major changes in the language of the speaker were required. Some care, however, in the use of the microphone, increase in articulatory precision, minor dialectal adjustments, etc., are reasonable to expect (Chao, 1956).

There are various ways in which the essential variables of the acoustical speech wave may be specified and quantified. Speech normally involves all of the elementary types of audible sound patterns. In particular, speech involves approximations to silence, periodic waves, random noise, and impulses. While the definition of these various forms is beyond the purpose of this paper, it should be noted that each can be specified.

Acoustical speech sound classes. Speech waves may be characterized by four types of acoustical time patterns which may be called acoustical speech sound classes. An

acoustical speech sound class is a class of time segments of speech involving a single acoustical form. Various analyses in the frequency and time domains may be performed in order to represent best the various parameters of these different types of acoustical waves. An *acoustical speech parameter* is a unidimensional time function which can be derived by means of a physical analysis of an acoustical speech sound class. The four acoustical speech sound classes and the parameters of which they are composed are as follows:

(a) *Quasi-periodic sounds* involving recurrent excitation by one or more vibrating mechanisms, plus resonance (and sometimes also anti-resonance) due to the source and transfer functions of the vocal cavities. An essential aspect of these sounds is the recurrent excitation (produced by the vocal cords, velum, tongue-tip, lips, etc.). In this acoustical speech sound class the spectrum and over-all amplitude may vary as a function of time. Parameters: *fundamental frequency* and the *resonance characteristics*, i.e. the amplitudes, bandwidths, and frequencies of the resonances and anti-resonances.

(b) *Quasi-random sounds* having a spectrum which is essentially continuous due to the frictional nature of the sound-generating mechanism involved in their production within the vocal cavities. Both the spectrum and over-all amplitude may change as a function of time. Parameters: the *resonance characteristics*.

(c) *Gaps* or silence preceding, between, or following speech sounds. Parameter: (zero level) *over-all instantaneous speech power*.

(d) *Impulses* following a gap which represent the release of an explosive or implosive sound (egressive or ingressive air). Parameter: (impulsive rise time and peak level) *over-all instantaneous speech power*.

Various superpositions of these four basic acoustical speech sound classes may occur, of course, as quasi-periodic sounds simultaneously with quasi-random sounds.

Acoustical parameters of speech

We must now consider the analysis of speech waves into those acoustical speech parameters which are of primary significance for linguistic interpretation. Fant (1956) has previously discussed the interdependence of certain of the parameters. Most of the parameters merit further study, both in regard to their specification by instrumental procedures and in relation to their linguistic interpretation. It may be noted that most of the parameters involve relatively low information rates (Flanagan, 1956). Fortunately, techniques which may actually increase the precision of measurement beyond that possible with Fourier amplitude methods are now being successfully developed in some laboratories.

Vowel and consonant parameters. The vowels and continuant consonants of speech are identified primarily by the resonance characteristics. These characteristics are evident in either quasi-periodic or quasi-random acoustical speech sound classes. There is evidence that fundamental voice frequency may also be of significance in the identification of vowels (Miller, 1953).

The relative formant amplitudes not only depend upon the resonance characteristics of the vocal tract, but also depend upon the input glottal spectrum and the radiating characteristics of the mouth and nose. In addition, the amplitude *vs.* frequency property of the acoustical environment in which the speech is produced (e.g. the room characteristic) may vary considerably. Thus there is reason to question whether formant amplitude information is particularly relevant in automatic speech recognition.

Both gaps and impulses are significant in the identification of plosive phonemes. Gaps are primarily associated with voiceless plosives; voiced plosives have low frequency energy present preceding the impulse, and thus do not normally involve gaps as defined here. The impulse rise times and peak levels both appear to be distinctive characteristics of plosive releases.

Prosodic parameters. As indicated in the previous paper, the three essential physiological prosodic parameters of speech are vowel and consonant duration, fundamental laryngeal frequency, and speech production power. The duration of vowels and consonants have discrete values, which may be plotted on a time base. Fundamental laryngeal frequency is directly represented in the harmonic structure of the narrow band analysis of speech, where it may be identified as fundamental voice frequency. Speech production power is an essential physiological speech parameter, but is related to acoustical speech power (as derived from the acoustical signal) in a complicated manner. The perception of vowel and consonant duration, fundamental laryngeal frequency, and speech production power are subjects which merit much further research.

Acoustical parameters for automatic speech recognition. On the basis of the above discussion a set of information-bearing acoustical speech parameters may now be given. The following is a tentative list of the *information-bearing acoustical parameters* (or briefly, the informational parameters) *of speech.*

(a) F_1 frequency of vowel or consonant first formant
(b) F_2 frequecy of vowel or consonant second formant
(c) F_3 frequency of vowel or consonant third formant
(d) F_{z1} frequency of consonant first anti-resonance
(e) F_{z2} frequency of consonant second anti-resonance
(f) F_0 fundamental voice frequency
(g) d duration of successive vowels and consonants
(h) π instantaneous speech power
(i) $\bar{\pi}$ average speech power

The parameters of primary significance in vowel and consonant identification may differ somewhat from one language to another, but probably these parameters contain most or all of the information of significance in the identification of the vowel and consonant phonemes and the prosodemes of English.

As is well known, the sound spectrogram provides a compact display of the resonance characteristics of speech sounds. While previous spectrographic methods have been imprecise in some respects, techniques are now being developed which may

provide the necessary exactness for speech analysis. For phoneme recognition it appears that a sound spectrogram-type of analysis would be most relevant. If formant amplitude information proves to be of importance to phonemic interpretations, then some type of three-dimensional spectrographic representation is necessary, in which amplitude is presented more accurately than on a binary scale. (The display, of course, may be assigned whatever form is most convenient for data processing; it seems improbable that an optical type of display will prove most useful.) Obviously, separate analyses must be performed on the individual acoustical parameters in items (f) to (i).

NORMALIZATION

It is well established that the physical data basic to both phonemes and prosodemes do not have fixed magnitudes (Pike, 1947; Peterson, 1952). An examination of sound spectrograms makes it immediately obvious, for example, that the same absolute magnitudes of formant frequency do not obtain for the speech of men, women, and children. In a similar manner, different time functions of fundamental voice frequency may convey essentially the same information, although the frequency ranges involved are markedly different (Abdalla, 1960). The same information may be communicated by time variations in the parameter of speech production power whether the over-all voice level is low or high. Thus relationships within the frequency domain for the resonance characteristics, and relationships within the time domain for the prosodies of speech are highly significant in their interpretation. Obviously, such relationships must be derived from actual parameter magnitudes.

The derivation of parameter relationships is equivalent to performing a normalization on the acoustical speech parameters after the initial analysis. The interpretation of acoustical speech parameters is here considered primarily to involve the identification of steady-state positions and controlled movements. Thus normalization of the parameters in the time domain appears unnecessary. A logarithmic spectrum analyzer provides an example of an unsatisfactory combination of parameter analysis and normalization. In a logarithmic analyzer, the filter width is proportional to the frequency of analysis. In such an instrument the lower and the higher formants are presented differently and both are presented in a manner which is not clearly or easily related to either speech production or speech perception.

As a first approximation, however, logarithmic scales (not filters) offer some advantage for speech parameter normalization. Where frequency ratios are essentially constant, logarithmic scales provide similar patterns, regardless of the absolute magnitudes; the patterns differ only by a constant factor of displacement. In some instances, it may also be desirable to normalize speech amplitudes. As a first approximation it may be appropriate to normalize both speech power and voice frequency measurements to logarithmic representations. Further study of the frequency relations among the speech of different speakers of identical dialects is particularly needed. Such a study should include data for the voices of men, women, and children. The scaling techniques

developed by Stevens (1960) should be applicable; basic data are also needed, however, on elementary signals having the parameters characteristic of speech.

If parameter relationships in frequency cannot be adequately specified, a possible solution to frequency normalization for different speakers (of the same dialect) would be to have each speaker read a standard passage into the automatic speech recognizer. From a previous knowledge of the phonemic composition of the passage, the machine could then determine an optimal frequency normalization for the input signal parameters for any given speaker.

Logical and statistical interpretations

The correspondence between sets of acoustical speech parameters and the phonemes and prosodemes of speech is highly complicated (Peterson, 1957). It is an assumption of the present paper that the relation between acoustical speech parameters and phonemes and prosodemes of any given language must be established largely on an empirical basis.

Normal conversational speech is far from an idealized sequence of independent steady-state positions and controlled movements. The linguistic interpretation of any particular speech parameter is often affected by the values of other parameters which are associated with it. Sequential intereffects in the time domain may also be considerable. For example, intereffects among vowel and consonant sequences have been described in detail in the phonetic literature where they are identified as assimilation, and in the phonemic literature where they are recognized as one basis of allophonic variation. The intereffects probably result from attempts to balance the requirement of intelligibility against minimal effort in speech production and perception (Zipf, 1949).

According to the above proposed procedures of analysis, the acoustical speech parameters will appear essentially in continuous or analogue form both preceding and following normalization. The next procedure in automatic speech recognition should be to perform measurements on these parameters. From these measurements phoneme and prosodeme candidates and their probabilities may be established. A simplified statement of the objective is to develop a set of mechanical rules for reading sound spectrograms. The problem is basically that of determining a set of logical and statistical procedures for interpreting the information-bearing acoustical parameters of speech. The probability associated with each measurement would be determined by the degree to which a given set of parameters or parameter sequences corresponds to a previously determined statistic for a specific phoneme or prosodeme.

Because of inexactness in speech production and the interference of internally and externally generated noise, clear and unambiguous phoneme and prosodeme identifications cannot in general be derived directly from the acoustical speech parameters. Rather, in the initial reduction of the acoustical data to phoneme and prosodeme sequences, a figure of probability should be associated with each individual identification. In the case of phonemes, this means that several different phonemes may be candidates for any given position within the sequence, and that each will have a

probability value associated with it. It should be noted that the sum of the probabilities of the various candidates for any given position within the phoneme sequence should have unity as a limit. For convenience in the subsequent discussion, phoneme and prosodeme probabilities based on the measurement of acoustical speech parameters will be called *acoustical probabilities*.

Obviously, the development of logical and statistical procedures for deriving a set of phoneme and prosodeme candidates from the acoustical speech parameters is both an important and a challenging research problem. A device which will perform such logical operations on the acoustical speech parameters obviously is not simply an analogue-to-digital converter, for the conversion process involves something more than the routine quantization of the input parameters. This type of transformation from the continuous to the discrete is essential in automatic speech recognition.

Linguistic storage

The inaccurate phoneme and prosodeme sequences as derived from the normalized acoustical speech parameters, perhaps broken only by the natural pauses of the speaker, may be satisfactory for many restricted applications of automatic speech recognition. For general purposes, however, precision in the automatic recognition of speech should be greatly enhanced by the use of stored information about the language involved (Fry and Denes, 1958). The acoustical phoneme and prosodeme probabilities may then be referred to the stored information for final decisions regarding the printed output.

Under any circumstances of automatic speech recognition, it is obvious that an inventory of the phonemes and prosodemes to be recognized must be stored. In a minimum type of automatic speech recognition procedure, probability information might be disregarded. In such a procedure the acoustical recognizers would yield only highest ranking phoneme selections at the output. Stored word possibilities might then be employed to correct and to separate the phoneme sequences resulting from the acoustical identifications.

Data regarding the frequency of occurrence of the stored linguistic units, however, should be of considerable value in interpreting the acoustical outputs. Linguistic probabilities will, of course, be of greatest value when there is the most uncertainty about the acoustical outputs, i.e. when the acoustical probabilities are universally low. When the degree of certainty is low, it seems clear that accuracy in recognition would be enhanced by such probabilities. For convenience in the subsequent discussion, frequencies of occurrence of the various linguistic units stored in the recognizer will be called *linguistic probabilities*.

Phonemes. Probably the most elementary linguistic statistics (or at least the most obvious) that might be considered for automatic speech recognition are the individual phoneme frequencies of occurrence. A study has been made of the degree of correlation among several existing frequency lists of English consonants (Wang and Crawford, 1960). The authors found a sufficiently high agreement among the various frequency

counts to suggest that the frequency distribution of the sounds is relatively stable for a wide variety of situations. Since there is considerable redundancy in English (Shannon, 1951), however, sequential probabilities should be of greater value than the probabilities of isolated elements. As the number of elements in the sequence is increased, the storage required increases approximately exponentially; also, in general, longer sequences will be less likely to occur. It should be noted that higher order phoneme sequence probabilities subsume those of lower order. As discussed in the next section (words), however, word probabilities should be of much greater value in automatic speech recognition than the probabilities of phoneme sequences derived without regard for word boundaries.

The phonemes of a language may be identified according to the physiological speech parameter values involved in the formation of the phones of the canonical allophones of those phonemes. Such a phonemicization for Midwestern American English is suggested in Fig. 2. The figure is constructed according to basic considerations outlined in the paper by Peterson and Harary (1961); such a chart provides a means of classifying the phonemes and prosodemes of a language. For each phoneme the symbol from the International Phonetic Alphabet has been selected which most closely represents the phones of the canonical allophone of that phoneme. In the consonant chart the voiceless consonants are entered in the first position, and the voiced ones in the second.

Variants of [m,n,l,r] occur in sequential positions similar to those normally occupied by vowels in Midwestern American English. They involve consonant and not vowel formations, however; for example, the posterior "edges" of the tongue have a very different shape for all positions of English [l] or [r] than for lateralized or retroflexed vowels. Under certain conditions of stress the coalescence of Midwestern American English *I err* and *ire* illustrates the strong phonetic similarity of the [r] formation in the position normally occupied by vowels and in that normally occupied by consonants. Thus it appears that the distinction indicated by the use of both /ɚ/ and /r/ in the transcription of Midwestern American English results from morphemic rather than from phonemic considerations. Thus, for example, according to the above discussion, *bird* and *butter* might be written phonemically as /brd/ and /bətr/, respectively.

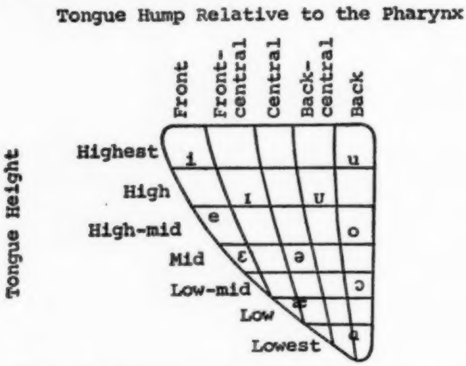
Words. It does not seem likely that the symbols for vowel and consonant phonemes, strung together continuously without spaces and punctuation, would be easily interpreted. The following sentence, in IPA notation, is an example of a phoneme sequence which might result from the direct printing of relatively accurate highest ranking identifications based on acoustical data:

ðɪszəsæmpləvðətɑɪpəskrɪpðətmɑɪtbɪɪkspɛktədwɪðəspitʃrɑɪtr

More complex linguistic units are clearly required to separate such sequences. While the linguists have found the "word" exceedingly difficult to define, such a definition seems unnecessary for the purposes of the present paper. One type of separation which should considerably aid the interpretation of the phoneme strings is the insertion of spaces between words. A stored vocabulary would, of course, be of considerable aid

VOWEL AND CONSONANT PARAMETERS

Primary Vowel Parameters



Primary Consonant Parameters

Place of Articulation

Manner of Articulation	Bilabial	Labio-dental	Dental	Alveolar	Alveolo-palatal	Palatal	Velar	Uvular	Pharyngeal	Glottal
Nasal	m			n			ŋ			
Sonorant	w			l	r	j				
Fricative	ʍ	fv	θð							h
Sibilant				sz	ʃʒ					
Trill										
Flap										
Plosive	pb			td			kg			
Click										

Modifying Vowel and Consonant Parameters

Air Direction
- Egressive - all
- Ingressive

Laryngeal Actions
- Whispered
- Breathy (V)
- Clear (V) - all
- Laryngealized (V)
- Voiceless (C) ʍ,f,θ,h,s,ʃ,p,t,k
- Voiced (C) m,n,ŋ,w,l,r,j,v,ð,z,ʒ,b,d,g

Secondary Articulations
- Spread (V)
- Rounded (V) u,ʊ,o,ɔ
- Labialized (C)
- Palatalized (C)
- Lateralized l
- Retroflexed r
- Velarized w,r,ʍ
- Nasalized
- Pharyngealized
- Glottalized (C)

PROSODIC PARAMETERS

Vowel and Consonant Duration
- Short Mid Long Extra-long

Fundamental Laryngeal Frequency
- Low Low-mid Mid Mid-high High Extra-high

Speech Production Power
- Weakest Weak Mid Strong Extra-strong

Fig. 2. The physiological parameter values of the phones of the canonical allophones of the phonemes of Midwestern American English.

in achieving such a separation, and should greatly facilitate the process of automatic speech recognition. With a large computer memory it should be possible to store most words which will occur, but new words are continually generated in language and not all proper names can be predicted. Obviously, the storage should be of those words which occur more frequently, and the frequency of occurrence may also be stored with each word. Since each word must be recognized in phonemic form, the words must be stored in the computer in phonemic form. In continuous speech some words may be represented by more than one phonemic sequence, however, and thus it may be desirable to store more than one phonemic pattern for some words. Obviously, it would also be possible to store conventional spelling with the individual words.

In most languages (including English) the permissible phoneme sequences within words are an essential aspect of the language structure. The storage of these allowable sequences might facilitate the separation of phoneme strings when words occur which are not contained in the computer memory. Since these would be unusual words, such as uncommon proper names, rare words, etc., it is questionable as to whether information about the frequencies of occurrence of the allowable sequences of phonemes would be particularly valuable. The frequencies of phoneme sequences within words will be different from the frequencies of those same sequences when tabulated without regard to word boundaries. Since an objective in automatic speech rcognition is to separate the phoneme strings into words, it appears that statistical data without regard to word boundaries would not be particularly useful in automatic speech recognition.

It is clear that the storage would be tremendously elaborated if word sequence probabilities were also incorporated. It does not seem likely that such probabilities would be of sufficient value in automatic speech recognition to justify the resulting complexity.

In addition to phonemic structure, stress is also an essential component of English lexicon. Thus it seems desirable in automatic speech recognition to store the normal stress pattern with each word. The properties of lexical stress in English have not yet been clearly defined. It has been demonstrated in various research studies that judgments of " lexical stress " in English are not only associated with speech power, but are perhaps even more strongly determined by fundamental laryngeal frequency and vowel duration (Fry, 1955, 1958 ; Lehiste and Peterson, 1959). The respective roles which the various prosodic parameters serve in lexical stress require much further quantitative study. The subdivision of vowel and consonant phoneme strings into lexical items should be considerably aided by such stress information. Since the measurement of stress from acoustical data, is highly complicated, however, the extent to which it will be of value in automatic speech recognition may be influenced by the complexity of instrumentation required for its specification.

Essential considerations in the use of lexical storage for automatic speech recognition include the relative number of words which cannot be anticipated (such as new words

or proper names) and the relative number of words which would cause ambiguities. With regard to ambiguities, there are three relationships among words which may be considered: words which are pronounced in the same manner and (a) are spelled the same but have different meanings (e.g. *seal, seal*); (b) are spelled differently and have the same meaning (e.g. *gray, grey*); and (c) are spelled differently and have different meanings (e.g. *stair, stare*). Only words of type (c) would directly cause ambiguities in the printed output of an automatic speech recognizer.

To obtain some initial estimate of the difficulties which the homonyms of type (c) would present, an examination was made of a list based on the 1,263 CNC monosyllables previously described by Lehiste and Peterson (1959). When proper names were excluded, 156 of the CNC words were homonyms which had at least two different spellings, and of these 156 words, 13 had three different spellings. While homonyms also occur among polysyllabic words, it seems reasonable to expect that the relative frequency of their occurrence will be appreciably less.

Lehiste (1959) has previously identified a number of acoustical cues to word boundaries. Whether such cues are sufficiently distinct and unambiguous and also sufficiently frequent to be useful in automatic speech recognition is yet to be determined.

An interesting complication to the separation of phoneme sequences on the basis of lexical information results from the fact that short words may be contained within longer words (e.g. *in, crease, increase*). Decisions about word spacings in such ambiguous cases can doubtless be made sometimes on the basis of stress; possibly other types of information may also be of value. In some instances, however, such ambiguities can only be resolved by the use of semantic information. The use of such information is, of course, beyond the scope of the procedures considered here. In these latter cases, in the absence of semantic information, it seems most reasonable to make the maximum number of intelligible lexical subdivisions.

Word groups. There is the additional problem of grouping words into larger linguistic units. Probably a semantic criterion would be the most difficult to implement. The use of semantic information would require processes very different from those discussed in this paper, and the possibility will not be considered here. The grouping of words into grammatical units, as sentences, is another possibility. Since an utterance may be meaningless but may have a normal grammatical form, the constraints on grammatical utterances are much less than are those on meaningful utterances. There are many different types of complete grammatical statements in a language such as English, however, and a very large number of words may normally be used in more than one grammatical category. Thus it would be exceedingly difficult to determine mechanically the possible grammaticalness of each potential word sequence. If the criterion of grammaticalness is imposed on the input, however, then stored syntactic information may be of practical value in automatic speech recognition. This possibility requires much further investigation and research.

Prosodemes. In addition to stored grammatical information, the acoustical speech waves contain information directly relevant to word groupings. In the acoustical waves

of speech such groupings are primarily indicated by variations in fundamental voice frequency, and in speech power (including pause or zero speech power).

Vowel and consonant duration, fundamental laryngeal frequency, and speech production power have been specified previously in physiological terms and have been called the physiological prosodic parameters of speech (Peterson and Harary, 1961). A prosody is a vowel or consonant duration or a prosodic parameter value or sequence of values which contains an approximation to a steady-state or a steady-state with an associated controlled movement. An alloprosody is the set of prosodies contained in the intersection of a maximal set of prosodically similar prosodies and a primary prosodically related set of prosodies. A prosodeme is the set of all alloprosodies which lie in primary prosodically related sets having non-similar prosodic environments, and which have canonical alloprosodies with pairwise minimal prosodic differences (not to exceed some specific upper limit). Prosodemes of vowel and consonant duration and fundamental laryngeal frequency are basic to linguistic systems of quantity and tone. Prosodemes of rhythm, intonation, stress, and accent appear to involve alloprosodies composed of various combinations of prosodies of vowel and consonant duration, fundamental laryngeal frequency, and speech production power. Further research on the prosodemes of individual languages is required in order to specify the manner in which such combinations are formed.

In the case of English, much has been written about the " suprasegmentals ", but as yet the prosodemes are not well defined. Intonation and lexical and syntactic stress appear to be the chief types. The identification of the various prosodemes of English, however, continues to be a major problem.

It is well known that prosodic information is not completely represented in the customary punctuation of English. Thus it is often possible to determine more precisely what is meant by hearing a message spoken than by reading the message in orthographic form ; the difference is due largely to the lack of detailed syntactic stress and intonational information in printed English. Thus prosodemes might be represented directly in the output of an automatic speech recognizer for English. The value of such information in printed material might be questioned, however, in view of the limited manner in which it is represented in the printing systems (punctuation) of many different languages. An alternate approach would be to reduce the prosodemic information to an approximation to conventional punctuation.

Trager and Smith (1951) have proposed that intonation contours in English are associated with various types of utterance terminations. While pause may be of assistance in identifying word groupings, pause is an aspect of speech style and may at times be a false cue to punctuation. There is an obvious need for further experimental investigation of the nature of intonation and stress and of their relation to syntactic units. In the automatic recognition of spoken English it appears that once identified, at least the most frequent stress and intonation prosodemes and their associated probabilities should be stored for reference.

Computational procedures

In the essential computer operation of automatic speech recognition, the phoneme and prosodeme candidates, based on the acoustical parameters, should be referred to stored linguistic probabilities. As indicated in the previous discussion, it seems undesirable to have the recognition procedures strongly influenced by linguistic probabilities when the acoustical probabilities are high. However, when the acoustical probabilities are low, recognition should be more successful if linguistic probabilities have a relatively larger influence upon the decisions. Thus the weighting given to the linguistic probabilities should depend upon the values of the acoustical probabilities. Further, it seems that the decisions should be a function of the degree of reliability of the acoustical analyzers. We may represent this reliability by a positive factor k; when k is large, the final decision should be more influenced by the acoustical than by the linguistic probabilities. Decision theory should provide the basic approach to the problem. However, as an illustration we may let:

$A_f \leq 1$ be the acoustical probability that the phoneme is f.
$L_f \leq 1$ be the linguistic probability that the phoneme is f.
k be a weighting coefficient.
P_f be the combined score derived for a particular phoneme f.

A simple procedure for combining the acoustical and linguistic probabilities may now be expressed by the relation:

$$P_f = kA_f + (1\text{-}A_f)\, L_f$$

where:

$$0 \leq P_f \leq k \text{ if } k \geqq 1$$
$$0 \leq P_f \leq 1 \text{ if } k \leq 1$$

The equation may be represented by families of straight lines with A_f as the independent variable, P_f as the dependent variable, and L_f as the parameter. The lines intersect at $P_f = k$ and the intercepts on the ordinate are L_f. When $k > 1$ all of the lines have a positive slope, and when $k < 1$ the lines for $L_f > k$ have a negative slope; the lines are parallel to the abscissa when $L_f = k$. In this relation the weighting given to the linguistic probabilities depends linearly upon the magnitude of the acoustical probabilities.

The resulting phoneme sequences may be grouped into words according to stored information about phoneme sequence probabilities and word occurrence frequencies. It may be possible to employ grammatical information in grouping the words into syntactic units. However, the prosodemes may provide the primary basis for syntactic decisions.

Print-out

As has already been discussed, the objective is to achieve the symbolic representation of the speech of a particular dialect, when produced under favourable acoustical conditions. As indicated above, in a general purpose automatic speech recognizer it should be possible to store the English spellings of most of the individual words to be printed in the output. Certain ambiguities will occur, as in the case of homonyms. Unusual words which cannot reasonably be anticipated for storage also present a problem. For such words some type of phonemic notation seems most appropriate.

Symbolization. Some of the problems in the choice of phonemic symbols for presenting the output of an automatic typewriter for English have been discussed by Fry and Denes (1957) and by Wang (1960). Symbols for transcribing the unusual words might be selected from the International Phonetic Alphabet according to the canonical allophones of the phonemes of the particular dialect involved. The transcription of such words might have a less traumatic effect upon the reader, however, if symbol selection were based upon conventional spelling.

When an alphabet is based upon conventions of English spelling, it is obvious that it cannot be a strict phonemic alphabet. It is common in English, for example, to represent a sequence of two phonemes by a single letter. A major criterion in the selection of the symbols suggested below is the frequency with which they represent various phonemes in conventional spelling.

Vowels. The vowels of English present more serious problems of phonemicization than the consonants, and phonetically they differ considerably from one dialect to another. Gleason (1955) has summarized the competing transcriptional systems proposed by various phoneticians and linguists. Many of the differences concern the duration and glide features (in contrast to transitions) associated with the vowels. A tentative vowel symbolization to be employed in transcribing unusual words in Midwestern American English is shown in Table 1. The symbols have been chosen to approximate common English spellings, rather than to preserve phonetic or phonemic conventions of transcription. One symbol, ə, has been introduced which is not found in conventional spelling. Since the sound is very common in unstressed positions in English it should be interpreted readily. The vowel portion of words such as " hose " and " late " appear to involve a sequence in which a steady-state is followed by a characteristic movement, or conversely. According to the previously mentioned paper, such sequences are treated as one phoneme, whereas diphthongs which involve two steady-state conditions are considered sequences of two phonemes. Since there is a common spelling for most of the diphthongs in English, these are also included in the proposed vowel orthography and are listed in Table 1. The previously mentioned IPA symbols for the phonemes and phoneme sequences are shown in the first column.

Consonants. There is rather good agreement among phoneticians and linguists regarding the consonant system of English. Also, the consonants do not differ greatly in their phonetic characteristics from one dialect of English to another. A tentative

TABLE 1

IPA SYMBOL	ORTHOGRAPHIC SYMBOL	KEY WORD
i	ee	f*ee*t
ɪ	i	s*i*t
e	ai	p*ai*n
ɛ	e	s*e*t
æ	a	s*a*t
ɑ	o	t*o*p
ɔ	au	t*au*ght
o	oa	g*oa*l
ʊ	u	f*u*ll
u	oo	t*oo*l
ʌ, ə	ə	s*u*ch, *a*bout
*ɑɪ	ei	h*ei*ght
*ɔɪ	oi	t*oi*l
*ɑʊ	ou	b*ou*t

Tentative list of vowels and diphthongs* with characteristic spellings for the automatic recognition of English.

TABLE 2

IPA SYMBOL	ORTHOGRAPHIC SYMBOL	KEY WORD	IPA SYMBOL	ORTHOGRAPHIC SYMBOL	KEY WORD
ʍ	wh	*wh*at	w	w	*w*att
f	f	*f*at	v	v	*v*at
θ	<u>th</u>	*th*ick	ð	th	*th*is
s	s	*s*at	z	z	*z*oo
ʃ	sh	*sh*ould	ʒ	zh	rou*ge*
*tʃ	ch	*ch*at	*dʒ	j	*j*aw
h	h	*h*at	j	y	*y*es
p	p	*p*at	b	b	*b*at
t	t	*t*ap	d	d	*d*ad
k	k	*k*it	ɡ	g	*g*ive
			m	m	*m*at
			n	n	*n*ot
			ŋ	ng	si*ng*
			l	l	*l*augh
			r	r	*r*est

Tentative list of consonants and affricates* with characteristic spellings for the automatic recognition of English.

list of consonant symbols for use in the automatic recognition of unusual words of Midwestern American English is shown in Table 2. In general, the orthographic symbols have been assigned according to the phonemes which they most frequently represent in ordinary English spelling. According to the previously mentioned theory, the affricates /tʃ/ and /dʒ/ would consist of two phonemes. Since they have a characteristic spelling in English, however, this has been preserved in the proposed system, as shown in Table 2. Symbols selected from IPA for the phonemes and phoneme sequences are shown in the columns at the left of the orthographic symbols.

Punctuation. The symbols of English punctuation may also be employed. It would not be simple, however, to follow the present conventions in their use. For example, some commas could be inserted, but it appears unlikely that the positions for colons and semi-colons could be distinguished consistently from those of commas without semantic information. Both periods and question marks could be inserted according to syntactic stress and intonation, but their resulting distribution would then differ from that found in conventional punctuation.

Instead of attempting to approximate conventional English punctuation, symbols might be inserted in the recognizer output for the direct representation of prosodemes of syntactic stress and intonation. A procedure for linearizing the combined phoneme and prosodeme representation of utterances has been outlined previously (Peterson and Harary, 1961).

Summary

Many would agree that a revision of English orthography is long overdue. It might be hoped that the revision would be hastened by the development of automatic speech recognition. Since conventional spellings can be stored with the phonemic forms of the more common words, however, it appears that most of the printed output of an automatic speech recognizer could be presented in conventional spelling.

We may now tabulate a general procedure for the automatic recognition of utterances within a particular dialect of a language. The essential operations are indicated in the schematic diagram of Fig. 3.

(a) The analysis of the input speech wave into a series of basic acoustical parameters.

(b) The normalization of the acoustical speech parameters into a standard frequency form for interpretation.

(c) The representation of the normalized acoustical parameters by a discrete set of phoneme and prosodeme candidates and their associated acoustical probabilities. The conversion may be based on an empirical set of logical operations performed upon the normalized acoustical speech parameters.

(d) Interpretation of intonation and stress prosodemes by referring the prosodeme candidates to stored information. These prosodemes may be represented in the output of the recognizer, or may be used as the basis for printing conventional punctuation marks.

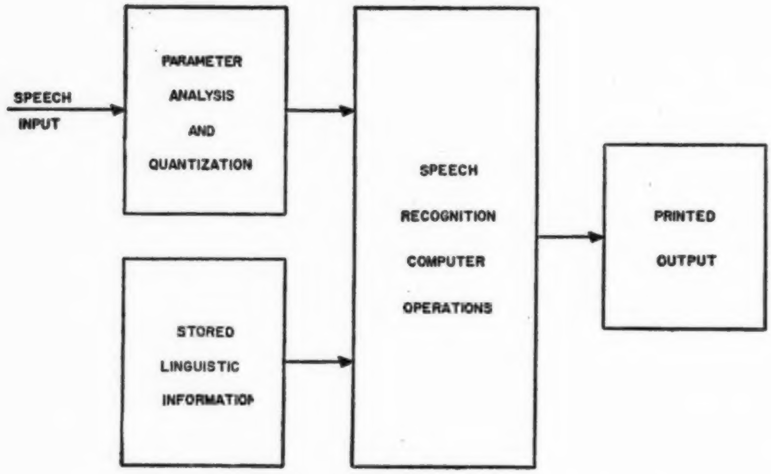


Fig. 3. Schematic diagram of automatic speech recognition procedures.

(e) Interpretation of the phoneme candidates and their probabilities and grouping them into words. The interpretations should be made by means of an optimization procedure, in which relatively long phoneme strings are analyzed concurrently; i.e. the data for each phoneme sequence should be stored and then interpreted with reference to the entire sequence. This process is very different from a serial procedure in which irrevocable decisions are made in a time ordered series.

(f) Grouping of the word sequences according to the prosodemes, and according to stored grammatical information when it can be applied.

(g) Print-out into words separated by spaces and grouped according to punctuation marks. Common words could be printed in ordinary English, but unusual words would require a special alphabet, such as that suggested in Tables 1 and 2. If desired, symbols for prosodemes of intonation and syntactic stress may also be printed. A system of punctuation, based on the symbols customarily employed in punctuating English, may instead be printed.

While much research has been conducted which is relevant to all of the above procedures, each continues to present major technical problems. The resolution of these problems requires continued basic research as well as instrumental development in automatic speech recognition. While the total system implied by the above procedures may now seem highly complex, this complexity does not appear to exceed the promise of the rapidly developing electronics and associated computer technology.

References

ABDALLA, A. G. (1960). An instrumental study of the intonation of Egyptian Colloquial Arabic. Ph.D. dissertation, University of Michigan.

CHAO, Y. R. (1956). Linguistic prerequisites for a speech writer. *J. acoust. Soc. Amer.*, **28**, 1107.

FANT, C. G. M. (1956). On the predictability of formant levels and spectrum envelope from formant frequencies. *For Roman Jakobson* ('S-Gravenhage), 109.

FANT, C. G. M. (1958). The acoustic theory of speech production. Royal Institute of Technology, Division of Telegraphy-Telephony, Stockholm, Report No. 10.

FLANAGAN, J. L. (1956). Band width and channel capacity necessary to transmit the formant information of speech. *J. acoust. Soc. Amer.*, **28**, 592.

FRY, D. B. (1955). Duration and intensity as physical correlates of linguistic stress. *J. acoust. Soc. Amer.*, **27**, 765.

FRY, D. B. (1958). Experiments in the perception of stress. *Language and Speech*, **1**, 126.

FRY, D. B. and DENES, P. (1957). On presenting the output of a mechanical speech recognizer. *J. acoust. Soc. Amer.*, **29**, 364.

FRY, D. B. and DENES, P. (1958). The solution of some fundamental problems in mechanical speech recognition. *Language and Speech*, **1**, 35.

GLEASON, H. A., JR. (1955). An Introduction to Descriptive Linguistics (New York).

LEHISTE, I. (1959). An acoustic-phonetic study of internal open juncture. Report No. 2 of the Speech Research Laboratory, University of Michigan.

LEHISTE, I. and PETERSON, G. E. (1959, a). Linguistic considerations in the study of speech intelligibility. *J. acoust. Soc. Amer.*, **31**, 280.

LEHISTE, I. and PETERSON, G. E. (1959, b). Vowel amplitude and phonemic stress in American English. *J. acoust. Soc. Amer.*, **31**, 428.

MILLER, R. L. (1953). Auditory tests with synthetic vowels. *J. acoust. Soc. Amer.* **25**, 114.

PETERSON, G. E. (1952). The information-bearing elements of speech. *J. acoust. Soc. Amer.*, **24**, 629.

PETERSON, G. E. (1957). The discrete and the continuous in the symbolization of language. *Studies Presented to Joshua Whatmough* ('S-Gravenhage), 209.

PETERSON, G. E. and HARARY, F. (1961). Foundations of phonemic theory. *Proceedings of the Symposia in Applied Mathematics.* American Mathematical Society (Providence, Rhode Island).

PIKE, K. L. (1947). The Intonation of American English (Ann Arbor).

SHANNON, C. E. (1951). Prediction and entropy of printed English. *Bell System Technical Journal*, **30**, 50.

STEVENS, S. S. (1960). The psychophysics of sensory function. *American Scientist*, **48**, 226.

TRAGER, G. L. and SMITH, H. L., JR. (1951). An outline of English structure, *Studies in Linguistics: Occasional Papers*, 3.

WANG, W. S-Y. (1960. Phonemic Theory A, with Application to Midwestern English. Ph.D. dissertation. University of Michigan.

WANG, W. S-Y. and CRAWFORD, J. C. (1960). Frequency studies of English consonants. *Language and Speech*, **3**, 131.

ZIPF, G. K. (1949). Human Behavior and the Principle of Least Effort (Cambridge).

CONTINUITY OF SPEECH UTTERANCE, ITS DETERMINANTS AND ITS SIGNIFICANCE

FRIEDA GOLDMAN-EISLER
University College, London

Pause frequency and word length of speech sequences uttered without break (referred to as "phrases") were measured, using speech produced under a variety of conditions. Several subjects were recorded in each condition. Individual differences as well as differences due to change of speech situation were highly significant. The significance of continuity in speech production is discussed in the light of the specific nature of these various differences.

INTRODUCTION

The fact that speaking is rarely continuous sound production, but consists of the utterance of sound sequences of varying lengths interrupted by pauses has been pointed out in previous work (Goldman-Eisler, 1956). That such pauses can take up a considerable proportion of the total speaking time has also been shown. No measurements, however, had so far been made of the lengths of the word sequences uttered continuously and of the frequency of the individual phrases alternating with these word sequences. The former is a function of the latter, and is expressed as a number of words per pause which is at the same time a measure of the frequency of pauses in relation to word production.

The sequences of words uttered without break will be referred to as "phrases" and interruptions of the vocal utterance of not less than 0.25 sec. are classified as "pauses"; breaks of less than 0.25 sec. are not counted because they might be due to changes of articulatory position and delays in the articulatory process as such.

TECHNIQUE OF MEASUREMENT AND MATERIAL

In order to measure the word length of the individual phrases sandwiched between pauses, visual transformations of speech have to be obtained, transferring the sound impulses to paper, using a pen-recorder (see Goldman-Eisler, 1956, 1958) and the visual tracings have to be synchronised with the verbal content of the records. This was done for the speech uttered in a previously reported experiment (Goldman-Eisler, 1961) in which nine subjects were asked to describe the events occurring in serial cartoon stories without captions (of the kind regularly published in the "New Yorker") and to formulate the meaning, point or moral of the story. They were also asked to repeat these descriptions and the summaries six times. Experimental conditions were thus created for the study of verbal behaviour (a) when speech is produced within a relatively concrete situation, i.e., a given sequence of events (through their description) and (b) in speech uttered in the process of abstracting and generalising from such

events (through summarising their meaning). Apart from the specific nature of the verbal task which stimulated this speech material, the situation was such that the speaker was not restricted in time and that his speech was a monologue.

In order to find out whether word length of phrases reflected the specific speech situation and nature of the verbal task the mean phrase length (in words) was also calculated for two basically different speech situations, namely discussions and psychiatric interviews which had been recorded in connection with previous experiments. In either of these speakers are constrained less and in a different way than in descriptions of pre-selected material. There is greater fluidity and scope for changing subject matter depending on the speaker's own initiative as well as that of his interlocutor. Furthermore, discussions are dialogues (or multilogues) and the time available to each speaker is always liable to be encroached on and cut short by the interlocutor.

While this is also true, to some extent, of the psychiatric interview, there are wide differences of purpose, discipline and relationship of interlocutors. In discussions, contestants are of equal status, while difference of status is the hall-mark of the interview situation. The benevolent and usually permissive superiority of the psychiatric interviewer aims, among other things, at relaxing the patient's sense of urgency and time pressure. His professional aim is to stimulate and guide the patient to talk freely about that which concerns him. The interviewer's utterances are geared mainly to this purpose, the interviewee is the acknowledged main speaker. In this sense the interviewee is more like the subject asked to describe and summarise cartoons. He too has the benefit of being allowed plenty of time in which to express himself.

For the discussions, no measurements of the individual phrase word length were available, but the mean phrase lengths were calculated by dividing the number of words produced by the number of pauses. Such data were gained for two groups of discussions:

(a) Discussions on controversial topics by academic personnel (4 subjects A,B,C,D ; one group discussion between A, B and C, and one discussion between C and D).

(b) Similar discussions in groups between adolescent boys drawn from middle-class and working-class backgrounds, and matched for verbal and non-verbal intelligence. (I am indebted for this material and the data underlying the measure of phrase length to Mr. Basil Bernstein.)

A wide range covering individual as well as group differences of class, education and intelligence was thus drawn upon to represent speech in the situation of discussion.

For psychiatric interviews measurements of the word lengths of individual phrases were available. The word length of individual phrases was also obtained for a further variant of speech situation, the speech being derived from the cartoon experiment after the descriptions of the cartoons and the summaries of their meanings had been repeated seven times. While in the original cartoon experiment as performed for the first time and in the discussions, speech was spontaneous, i.e. organised anew, at the time of speaking, in the repetitions speech had become more or less automatic, the result of practice, the verbalisation of well-learned connections.

TABLE 1

Situation : CARTOON EXPERIMENT

Verbal Task :	Describing cartoons		Summarising meaning	
	1st time	*7th time*	*1st time*	*7th time*
Subjects	W/P	W/P	W/P	W/P
Tho	6.0	8.8	4.6	4.0
Ha	4.4	7.2	3.6	6.1
Tr	4.3	5.4	3.8	4.9
Wi	3.7	8.3	3.9	10.9
Sa	4.6	5.3	5.5	6.5
Gi	5.6	6.0	5.6	4.2
Ne	5.5	7.0	5.7	5.7
Aa	3.2	3.7	2.5	3.3
Do	4.7	7.1	5.2	5.9

Situation : DISCUSSION
Verbal Task : Debating

Subjects :	(1) Adult academic workers	(2) Adolescents	
	W/P	W/P	
A	5.3	11.7	6.3
B	11.1	9.3	7.2
C	12.9	4.3	2.8
C	5.8	8.0	11.1
D	11.1	5.6	8.7
		5.6	8.2
		8.0	7.2
		6.8	8.2

Situation : PSYCHIATRIC INTERVIEW
Verbal Task : Relating personal history and background of illness.

Subjects :	Neurotics
	W/P
M	5.3
B	6.3
S	6.3
H	6.4
P	5.2

Phrase length, i.e. number of words per pause (W/P): means for individuals.

RESULTS

(1) Table 1 shows the mean phrase lengths measured in number of words of individuals in the different speech situations. The figures for descriptions of cartoon series and the formulation of their meaning done for the first time and for the seventh time are drawn from the same nine subjects (9,743 words). The figures for the discussions are drawn from a group of different subjects, twenty-one altogether (9,950

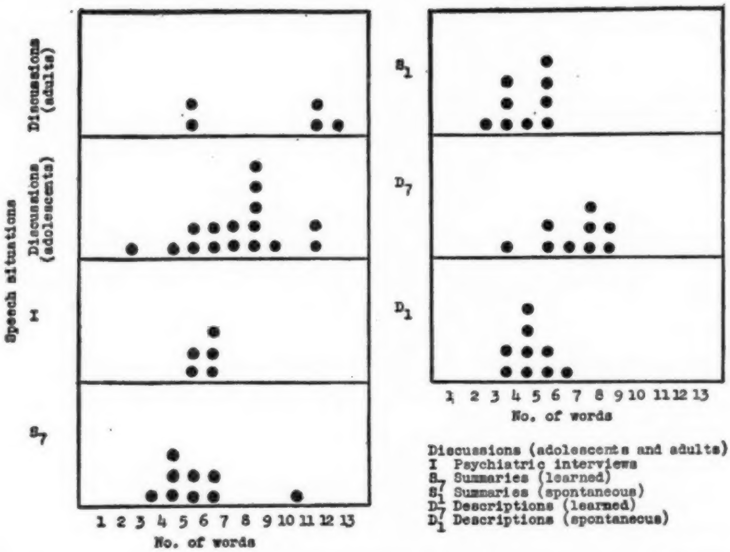


Fig. 1. Frequency distributions of mean phrase lengths for different speech situations.

words). The figures for the psychiatric interviews are based on 5,096 words uttered by five patients. As may be seen from the distributions in Fig. 1, the discussions cover the widest range of phrase length. Their distribution embraces the distributions of the same measure derived from the speech uttered when describing cartoons and formulating their meaning for the first and seventh time, and those derived from the interviews.

(2) *Cartoon Descriptions and Summaries.*

An analysis of variance was computed for the data derived from the cartoon experiment comparing (a) the individual speakers, (b) descriptions with summaries, and (c) the original and spontaneous versions with the well-learned repetitions. We can see from Table 2 that individual differences are highly significant ($p < 0{\cdot}001$), and that differences in the conditions of speech production, such as between spontaneous utterance which is coincident with the verbal planning on the one hand and the utterance of familiar, well-learned sequences after several repetitions on the other, have a systematic effect on the phrase length of individuals ($p < 0{\cdot}001$). Well-learned sequences have greater continuity, repetition leading to a closure of gaps, particularly in descriptive speech where the mean increase of continuity was from 4·8 to 7·4 words. Fluency in the utterance of summaries is enhanced to a much lesser extent (from 5·2 to 6·5 words). At the same time the difference of level of verbal planning which distinguishes the description of concrete events from the formulation

TABLE 2

SOURCE	SUM OF SQUARES	*df*	VARIANCE ESTIMATE
Descriptions and summaries	4.36	1	4.36
Before and after learning	316.05	1	316.05
Between subjects	538.71	8	67.35
Interaction	466.00	305	58.25
Error	2010.33	8	6.59
		323	

F Descriptions/Summaries/Error	= 0.66 N.S.	
F Before and after learning/Error	= 47.98	$p < 0.001$
F Subjects/Error	= 10.17	$p < 0.001$
F Interaction/Error	= 8.84	$p < 0.001$

Analysis of Variance.

Phrase length (W/P) for descriptions and summaries uttered before and after learning, based on 9 subjects and 9 cartoons.

of their meaning and which was so clearly reflected in the length of pauses (Goldman-Eisler, 1961), is not reflected in degree of continuity as measured by lengths of phrases (or frequency of pauses); it is, however, reflected in the gain of fluency after learning has taken place.

The interaction between individuals and effect of learning on continuity of speech production is highly significant. This interaction is particularly significant in the formulation of summaries, and F ratio for these being 3·30 ($p = 0{\cdot}001$) while for descriptions it is 2·77 ($p < 0{\cdot}01$). In other words, the effect of repetition on continuity depends on the individual more in the summaries than in the descriptions (Table 3). This is further substantiated by the results of the analysis of variance performed on the gains (or losses) in continuity after learning (Table 4). This showed that the discrepancy between first and seventh version (gain or loss in phrase length) was a matter of the individual speaker to a much greater extent in the summaries than in the descriptions. The F ratio comparing the variance for individuals with the error variance was 4·44, $p < 0{\cdot}001$ for the summaries, but only 2·60, $p < 0{\cdot}05$ for the descriptions.

Examining the data, it became clear that different individuals responded to repetitions differently, some gaining in continuity whilst others showed greater discontinuity in the summaries after learning than before. Descriptions, on the other hand, were generally uttered with considerably greater fluency after learning.

Learning thus has a greater levelling effect on individual differences when speakers respond to the easier verbal task of describing events than in the more difficult operation of formulating their meaning. This is further substantiated by the differences in F ratio when comparing the phrase length of learned with spontaneous speech separately

TABLE 3a

SOURCE	SUM OF SQUARES	*df*	VARIANCE ESTIMATE
Between subjects	304.33	8	38.04
Before and after learning (blocks)	282.43	1	282.43
Interaction (block and subjects)	121.21	8	15.15
Error	789.17	144	5.48
		161	

F Learning/Error $= 51.53$ $p < 0.001$
F Subjects/Error $= 6.94$ $p < 0.001$
F Before and after learning/Subjects $= 7.42$ $p < 0.05$
F Interaction/Error $= 2.77$ $p < 0.01$

Analysis of Variance.

Phrase length (W/P) for descriptions uttered before and after learning, based on 9 subjects and 9 cartoons.

TABLE 3b

SOURCE	SUM OF SQUARES	*df*	VARIANCE ESTIMATE
Between subjects	318.90	8	39.86
Before and after learning	69.47	1	69.47
Interaction (block and subjects)	224.40	8	28.05
Error	1221.16	144	8.48
		161	

F Learning/Error $= 8.19$ $p < 0.01$
F Subjects/Error $= 4.70$ $p < 0.001$
F Interaction/Error $= 3.30$ $p < 0.01$ (nearly 0.001)

Analysis of Variance.

Phrase length (W/P) for summaries uttered before and after learning, based on 9 subjects and 9 cartoons.

for descriptions and summaries. The F ratio for the former was 51·5, $p < 0{\cdot}001$ and for the summaries 8·2, $p < 0{\cdot}01$ (see Table 3).

(3) The mean phrase length of the individuals performing the verbal tasks of describing and summarising the meaning of pictures for the first time ranged between 3·2 and 6·0 words per phrase (i.e. also per pause) for the descriptions and between 2·5 and 5·7 words for summaries. This does not, however, mean that phrases of greater length or shorter did not occur. The above figures are averages covering distributions of an extremely wide range, but of great skewness with the mode in the very low values of one to three words per phrase according to the individual. The maximum phrase length reaches 23 words.

TABLE 4

DESCRIPTIONS

SOURCE	SUM OF SQUARES	*df*	VARIANCE ESTIMATE
Between subjects	226.53	8	28.32
Between cartoons	81.22	8	10.15
Error	648.47	64	10.13
		80	

F Subjects/Error = 2.60 $p < 0.05$
F Cartoons/Error = 1.00 N.S.

SUMMARIES

SOURCE	SUM OF SQUARES	*df*	VARIANCE ESTIMATE
Between subjects	463.40	8	57.93
Between cartoons	169.17	8	21.15
Error	835.68	64	13.05
		80	

F Subjects/Error = 4.44 $p < 0.001$
F Cartoons/Error = 1.62 N.S.

Analysis of Variance.

Effect of learning on phrase length (W/P), based on 9 subjects and 9 cartoons.

Fig. 2 shows the cumulative percentage distribution of the mean phrase lengths for descriptions and summaries, uttered for the first and seventh time. It shows that 50% of speech is broken up into phrases of less than three words, 75% into phrases of less than five words, 80% into less than six words, 90% less than ten words, and that phrases of more than ten words uttered with fluency constitute only 10% of speech when speakers are engaged in describing pictures for the first time.

After having repeated their performance six times (in some cases seven, eight or nine times) only 35% of speech is broken into phrases less than three words, while 50% are less than five words in length, 65% less than six words, 85% less than ten words, and 90% less than twelve words. Thus even in speech as well learned as this, phrases with more than ten words are uttered in only 15% of cases.

The cumulative percentage distribution of the mean phrase lengths for psychiatric interviews is practically identical as regards the middle values with this same distribution for descriptions after learning, but has a somewhat smaller proportion of very short phrases (less than three words) and contains more very long phrases, going up to 30 words (see Fig. 2). Of the five speech situations analysed for individual phrases, interviews showed the highest continuity, and we may assume from the mean values that discussions show even greater continuity.

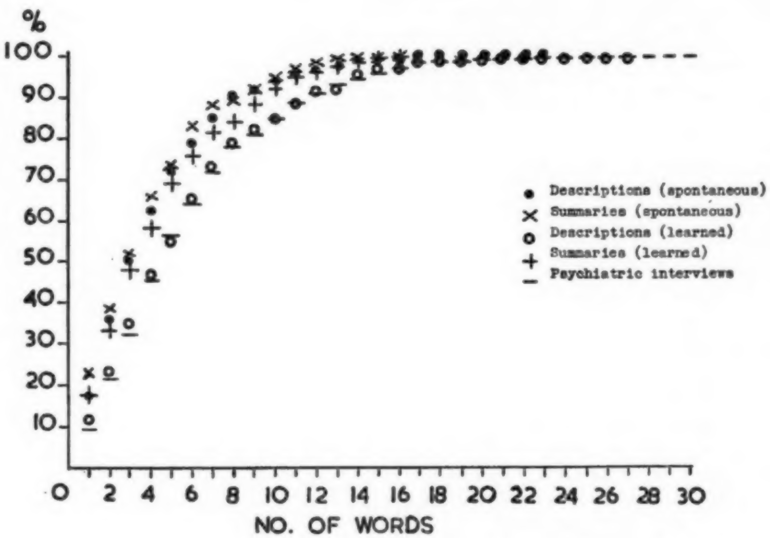


Fig. 2. Cumulative frequency distributions (percentages) of phrase lengths in five speech situations.

It should be noted that the criterion of continuity or fluency is a lenient one, the minimum break classified as a pause being 0·25 sec.; in other words, allowance is made for gaps of up to 0·25 sec. which occur much more frequently than longer breaks and which may serve articulatory or rhetorical purposes. Thus fluency in our sense is not necessarily a complete continuity of sound production, but continuity within such limits as might be thought necessary for well articulated, structured and intelligible speech.

Discussion and conclusions

The significance of phrase length and its dependence on the situation in which speech is uttered might be better understood if we consider the following facts:

The mean phrase lengths of speech recorded in discussions was 7·5 words, which is the same as the phrase lengths of highly practised descriptions (7·5 words) as against 4·8 words in the original descriptions. Before drawing further conclusions it should be noted that the speakers were different people, but that the sample in the discussions was larger (21 speakers) and more widely distributed socially and educationally. If we single out the adult group socially and educationally corresponding to the speakers in the descriptions (cartoon experiment) we observe a very wide distribution of phrase lengths covering the same range as those based on the sample of adolescents. A further

indication may be gained from the fact that one of the speakers in the discussions of the academic adult group was also a subject in the cartoon experiment and that his mean phrase length in discussion was 11·1 words while it was 5·9 words in the descriptions, and 9·2 words after seven repetitions. A further datum indicative in this context is that speaker " C " of the adult group of whom we have figures for two different discussions, produced a phrase length of 12·9 words on one occasion and of 5·8 words on the other. This seems to be in contradiction to the fact of individual consistency unless a basic difference in the conditions of speech production can be shown to have existed which might be responsible for such a shift, one which is comparable to the shift from spontaneous to well-learned descriptions. On examining the context, such a difference was indeed found.

In the first discussion (phrase length = 12·9 words) " C " talked to her colleagues treating the debate in a joking and light-hearted manner. The second discussion (phrase length = 5·8 words) in which " C " debated against her chief, an extremely skilful debater, was a situation of much greater challenge and to demonstrate a high quality of argument was for " C " a matter of some consequence.

Obviously discussions allow for a greater generality of conditions than situations in which the subject is directed to devote himself to one special operation, and taking the above facts into account, one might argue as follows: Discussions are a type of situation which allows for a mixed bag of operations. Automatic verbalisation of well-learned sequences will alternate with the utterance of words and expressions individually selected and fitted to the occasion, with the new formulation of general statements, etc. The proportion of each type of speech, and level of speech planning will depend on the type of discussion, the demands made on the speaker either by interlocutor, or theme, or other factors in the situation (e.g., listeners), the speaker's individual disposition, or the time factor inherent in the situation and the extent to which it is inimical to delay. If phrase length is an indication of how automatic the utterance of a sequence has become, as we may now be justified in assuming, speakers in discussions and conversations generally are more fluent than in the verbal tasks of describing and summarising the meanings of events perceived when no time limit was imposed, because there is either no need, or under the pressure of the interlocutor's possible interruption, little encouragement to indulge in or rise to reflective speech. Only when intellectual quality is at a particular premium, as was the case in the discussion " C " vs. " D ", the motivation to reflect was stronger than that to keep talking. Fluency had to be sacrificed to intellectual quality—to good rather than glib argument.

The greater fluency in general of discussions is thus a reflection of the fact that the situations of social interaction are not as conducive to verbal planning of a high quality as those in which speakers are allowed to soliloquize.

Further evidence confirming this conclusion comes from the measurements of the phrase lengths in psychiatric interviews. Fig. 2 shows that the cumulative percentage distribution of frequencies of the mean phrase lengths for the five interviews analysed coincides in the main part with those of the descriptions after learning, and differs from the spontaneous descriptions even more than from the former inasmuch as it

contains, at the tails of the distribution, even longer phrases at the one end, and a smaller number of one-, two- and three-word phrases at the other. Psychiatric interviews are devoted to the recall of emotionally charged material from patients' personal histories. The emphasis is on the production of the raw material of experience. The more successful the interview, the less selective is the interviewee towards his material and the way he presents it and the more he follows association by contiguity. Hughlings Jackson observed (1932) that emotional speech is a sub-class of automatic speech. The distribution of phrase lengths in psychiatric interviews is in keeping with this classification.

The relevance of level of intellectual activity to phrase length is also exemplified by the fact that the levelling influence of learning on individual differences shows itself more clearly in the descriptions than in the summaries. The transformation of spontaneous, newly organised speech into well-learned automatic speech productions is more efficient when the verbal task is such that a lesser degree of abstraction is involved.

In the former, less difficult operation, i.e. in the descriptions, the verbal form achieved seems to be accepted with, or soon after, the first formulation, leaving scope for the learning process to achieve smooth utterance. The initial formulation is subjected to learning with confidence. Poor efficiency in the learning process would indicate that such confidence is lacking; that, in fact, learning may not be taking place, or only partially, because formulations are not accepted as final and are further elaborated.

The persistence of individual differences in spite of the same amount of practice in descriptions and summaries points to a residue of creative speech processes resisting automatic performance in some speakers more than in others. It indicates that some speakers tend to be less confident in accepting their first formulations as final than others, and the fact that this difference is much more pronounced in the summaries than in the descriptions points to a special significance of efficiency in the learning of spontaneous speech material as measured by increase in continuity of utterance. Efficiency in this sense, it would seem, is a function of the attitude of the learner to his material. Its degree is a measure of the extent to which speech utterance has ceased to be a creative process and its production a conscious effort. Efficiency in learning would be at its best when the choices made at the first attempt are accepted as final and at its worst when uncertainty lingers on and new choices are made or contemplated. The higher the intellectual level of verbal production the more the efficiency in transforming them into smooth utterance distinguishes between individuals.

Further information as to what kind of individual achieves smoothness of utterance more or less efficiently may also contribute towards our understanding of what is implied in the efficient transformation of choice into automatic speech action. To this purpose gain in continuity as measured by the increase in "phrase" length when formulating the descriptions and summaries for the first time was correlated with two hesitation phenomena, silent pauses and filled pauses. The former was measured in pause time per word produced (P/W) and the latter by the rate of occurrence of "ah" and "hm" sounds (or filled pauses) per second of silent (or unfilled) pause.

TABLE 5

	EL(d)	EL(s)	P/W(d)	P/W(s)	FP/UP(d)	FP/UP(s)
EL(d)						
EL(s)	0.617*					
P/W(d)	−0.208	0.279				
P/W(s)	−0.133	0.817**	0.703*			
FP/UP(d)	−0.267	−0.767*	0.036	−0.550		
FP/UP(s)	−0.767*	−0.667*	−0.600*	−0.633*	0.867**	

** Significant at 1% level.
* Significant at 5% level.

Intercorrelation table (Spearman's rank correlation coefficient) showing the effect of learning (EL).

EL(d) = effect of learning in descriptions.
EL(s) = effect of learning in summaries.
P/W(d) = pause time per word in descriptions.
P/W(s) = pause time per word in summaries.
FP/UP(d) = ratio of filled pause to unfilled pause time in descriptions.
FP/UP(s) = ratio of filled pause time to unfilled pause time in summaries.

Both these hesitation phenomena had previously (Goldman-Eisler, 1961a and b) been shown to be speech habits characteristic of individuals while at the same time reflecting different internal processes. Silent hesitation, i.e. an arrest of external vocal activity (speech or non-linguistic vocal action) was shown to reflect cognitive activity such as selection, abstraction or planning in speech, for periods proportional to the difficulty of the cognitive task, while vocal hesitation (" ah ", " hm ", " er " sounds) seemed to reflect emotional attitudes. The former, the silent hesitations distinguished sharply between descriptions and summaries; the latter, the vocal hesitations, were insensitive to the differences in the cognitive processes stimulated in these two verbal tasks.

Table 5 shows the correlations between the means for individuals of three speech parameters, (1) the gain in continuity through learning, (2) pause time per word produced, and (3) rate of occurrence of filled pauses per second of unfilled pauses. From this we see that the individuals who are most efficient in achieving continuity after frequent repetition in the final utterance of summaries were those who paused longest when formulating them for the first time. (There is no such relationship for the descriptions.) These speakers were also least given to vocal hesitations, i.e. to " ah "-ing and " hm "-ing.

We have previously (Goldman-Eisler, 1961 b) concluded that speakers consistently inclined towards delay of action and tolerance of silence achieved superior (i.e. more concise) stylistic and less predictable linguistic formulations, whilst those inclined towards greater verbal as well as vocal activity and to intolerance of silence were marked by inferior stylistic achievement (long-winded statement) of greater predictability. We may now add that the former achieve greater efficiency as measured by

smoothness or continuity of utterance when subjecting their productions to the process of learning.

In other words, initial delay in the production of speech accompanying verbal planning at a high level of cognitive activity, such as abstraction and generalisation, pays off in the ultimate efficiency of the process of reproduction.

The general maxim that time invested in reflection initially is compensated by greater efficiency ultimately has thus been shown to be valid in the field of verbal planning, and speech parameters have been evolved to describe the relevant relationships in quantitative terms.

References

GOLDMAN-EISLER, F. (1956). The determinants of the rate of speech output and their mutual relations. *J. Psychosom. Res.*, 1, 137.

GOLDMAN-EISLER, F. (1958). Speech production and predictability of words in context and the length of pauses in speech. *Language and Speech*, 1, 96.

GOLDMAN-EISLER, F. (1961a). Hesitation and information in speech. In *Information Theory: Proceedings of the 4th London Symposium on Information Theory* (London), 162.

GOLDMAN-EISLER, F. (1961b). A comparative study of two hesitation phenomena. *Language and Speech*, 4, 18.

JACKSON, J. H. (1932). Selected Writings, vol. II (London).

THE DISTRIBUTION OF PAUSE DURATIONS IN SPEECH

FRIEDA GOLDMAN-EISLER
University College, London

The subject of this paper is the duration of individual pauses and their distribution over extensive tracts of speech uttered in a variety of situations. The determinants isolated include individuals and conditions of speech utterance (social, emotional, cognitive). The analysis of pause time into duration of individual pauses and pause frequency was shown to be more powerful in gauging the type of decision involved in the production of speech than either of these parameters alone.

The main concern of previous studies which have centred around the measure of pause duration has been its relation to processes of selection, verbal planning and the generation of information in speech. (Goldman-Eisler, 1958, 1961.) The distribution of the individual pauses had so far not been studied. The measures used were the summated lengths of the individual pauses which occurred within the utterances (word sequences) concerned.

The present paper is concerned with the lengths of the individual pauses and their distribution over extensive tracts of speech. The material used for measurement was the same as that described in the previous paper (Goldman-Eisler, 1961).

The technique rests as before on the transformation of sound to visible recordings. The pen-recorder connected with the tape recorder was run at a speed such that 3 mm. on the visual record represented one second. The material is summarised in Table 1 where Description 1 and Summary 1 refer to the first spontaneous attempt at describing and summarising the meaning of cartoon stories from the " New Yorker " as reported previously (Goldman-Eisler, 1961) and Description 7 and Summary 7 refer to the seventh repetition of the spontaneous version (see also page 220 of this issue).

RESULTS

The following facts have emerged from these measurements:—

(1) The lengths of individual pauses are distributed differently for different individuals. This was shown by testing the significance of the differences between the frequencies of nine distributions of pause lengths derived from the nine subjects in the cartoon experiment. The χ^2 for $N = 32$ was 387·39, the probability being far less than 0·001. The frequencies were in classes of 0·5 sec. and for the data obtained through the cartoon experiment the frequencies in the class of less than 0·5 sec. were divided into those of less than 0·25 sec. and 0·25 to 0·5 sec. In previous work, gaps in speech of less than 0·25 sec. have been excluded from consideration under the term hesitation pause, so that this division has qualitative significance. As the ratio of the very short pauses—less than 0·25.—which we might call articulatory pauses, to the shortest

TABLE 1

SPEECH SITUATIONS	NO. OF SUBJECTS	NO. OF PAUSES	
Cartoon Experiment (Description 1)	9	1,367	
Cartoon Experiment (Summary 1)	9	694	
Cartoon Experiment (Description 7)	9	971	
Cartoon Experiment (Summary 7)	9	519	
Discussions (Adult)	5	529	943
Discussions (Adolescents)	8	414	
Psychiatric Interviews	7	969	

TABLE 2

SPEECH SITUATIONS	*Less than 0.5 sec.*	*1.0 sec.*	*2.0 sec.*	*3.0 sec.*	*3.0 - 8.0 sec.*	*8.0 sec. and over*
Cartoon Descriptions (spontaneous)	47.8	23.7	17.2	6.0	4.6	0.7
Cartoon Summaries (spontaneous)	43.6	19.8	16.3	8.8	9.6	1.9
Cartoon Descriptions (learned)	59.6	24.3	12.7	2.7	0.7	0.0
Cartoon Summaries (learned)	63.7	20.0	13.5	2.0	0.8	0.0
Discussions (Adults)	49.9	37.1	12.0	1.0	0.0	0.0
Discussions (Adolescents)	41.4	41.1	16.0	1.3	0.1	0.0
Psychiatric Interviews	16.4	33.9	28.6	10.8	9.6	0.6

Percentage occurrence (means) of pauses of different duration.

hesitation pauses proper (0·25 to 0·5 sec.) is about 2 to 1 in all these four speech samples, we might not be far out if we assume such a ratio in respect of the other speech situations.

(2) The distribution of pause lengths is determined by the type of situation in which speech is uttered. The cumulative frequency distributions based on the mean frequencies for each of the speech situations measured show considerable differences which are highly significant ($p < 0{\cdot}001$; see Fig. 1).

(3) The most remarkable fact about these differences (see also Table 2) is that pauses in discussions, whether among academic adults, or adolescent middle and working class schoolboys are never longer than three seconds and that 99% are less than two seconds.

(4) As with phrase lengths (Goldman-Eisler, 1961b) the distribution of pause lengths in discussions approaches that of the well-learned reproductions of the descriptions and summaries of cartoons more closely than any other, while spontaneous, first-time descriptions and, to an even greater degree, summaries occupy a distinctly different universe containing a smaller proportion of short and a larger proportion of long

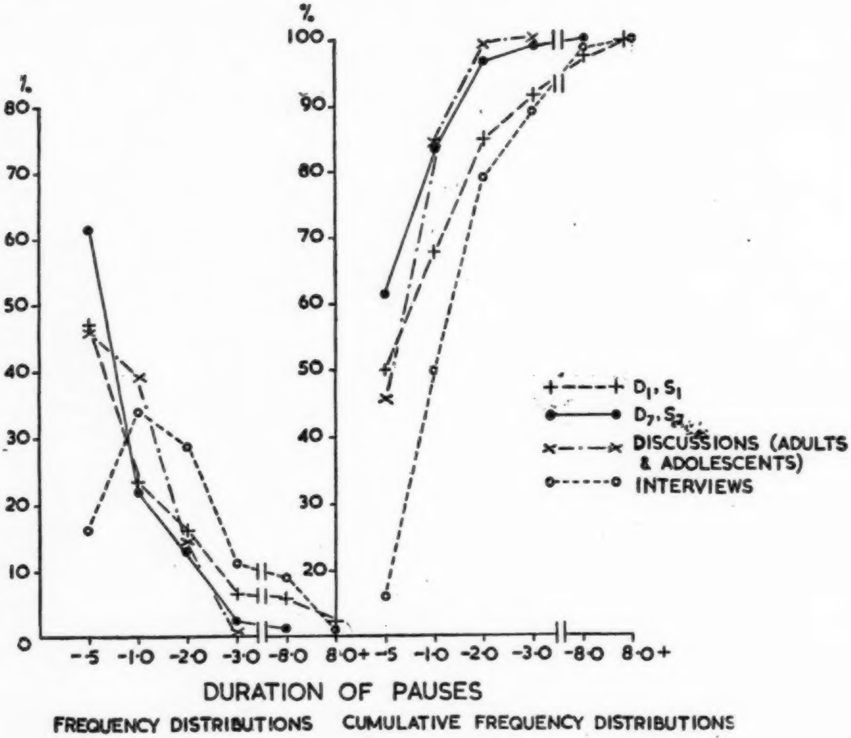


Fig. 1.

pauses. While 87·0% and 82·5% of all pauses in discussions are less than one second the proportion for the speech describing and summarising cartoon stories for the first time was only 71·5% and 65·9% respectively. In the latter we find sizeable quantities of pauses longer than 3 sec. (5·4% for descriptions and 10·4% for summaries) tailing off towards spans of more than 20 sec., whilst the discussion situation does not seem to permit at all a break in vocal production of more than 3 sec.

(5) The majority of the pauses in excess of 3 sec. (5·1% for descriptions and 8·1% for summaries) extend as far as 8 sec., but a small proportion (0·5% and 2·3%) of pauses can be very long indeed, the maximum being 30 sec.

(6) Psychiatric interviews have a smaller proportion of short pauses (50% less than one second) and a large proportion of pauses of middle length (up to 3 sec.), but have less of the extremely long pauses which characterize speech in the cartoon experiment.

In fact, the psychiatric interview was the only speech situation investigated in which pause lengths maintained a central tendency; pause length for the other conditions of speech was distributed exponentially, the frequencies fast diminishing after a pause length of one second but having a significant tail of very long pauses when descriptions and particularly summaries of the meaning of a cartoon were formulated anew.

Discussion

We have so far taken measurements of pauses in speech in three ways: (1) Total duration of pausing for any utterance, a quantity which may be rendered independent of the length of utterances by expressing it as a ratio of pause time to number of words (P/W in seconds), i.e. pause time to speech time ratio.
(2) Frequency of pausing, which, measured in relation to words produced, may be expressed as a ratio of one pause to so many words, or reciprocally as number of words produced per pause (W/P in words), which is also a measure of continuity, i.e. length of word sequence uttered without break (" phrase length ") and
(3) the duration of individual pauses as they alternate with individual word sequences. We may study their distribution in any speech sample investigated and derive from it a central value.

While the total pause time must be a function of the frequency and length of the individual pauses, the two components may be subject to variation, which would be in inverse direction. The function is hyperbolic ($y = ab$, where y is the total pause time and a and b are frequency and duration of individual pauses).

As for any particular total pause time a wide range of combinations of frequency and length of individual pauses from which it can result is conceivable, a new dimension of measurements is attained by taking the two factors into account and thus offers a wider scope for differentiating individuals and the conditions (psychological, physiological and social) of speech production.

Obviously information about any one of these parameters only is incomplete; for speech samples comparable on the basis of total pause time may be not at all comparable when frequency and length of individual pauses are taken into account separately. Equally, speech matched on the basis of frequency of pauses (and " phrase length ") may differ greatly when the length of the individual pauses is taken into account. And similarly, individual pauses of similar length may divide phrases of different length under different conditions or with different individuals.

This is indeed the case when we compare the distributions of pause durations for interviews with the other speech situations. For pause duration, interviews show some overlap with the original descriptions and summaries, the other speech situations taking up an entirely distinct space (see Fig. 1). As regards frequency of pauses and therefore " phrase length " or continuity on the other hand, the original descriptions and summaries (especially the latter) are furthest removed from the interviews; these are nearly identical with the frequency distributions of the well-learned reproductions of

the description (D 7). In other words, interviews approach in their proportion of long pauses the most intellectual speech production, i.e. those requiring the highest level of verbal planning, while in frequency of pauses they are comparable to the most automatic speech products requiring least verbal planning.

In the assembly of speech situations considered in this study the psychiatric interviews were unique in that they were the only ones whose function was to stimulate the expression of emotion by way of recalling emotionally charged material from personal history. The rest were dedicated either to the display of cognitive functions such as those required in debating controversial issues (as in discussions) or in the verbal encoding of pictorial contents and the re-coding of their meaning, i.e. the display of powers of symbolic expression, or to the display of processes of learning and reproduction.

It is therefore interesting to see that it is in this condition of speech production, namely the psychiatric interview, and only this, that we find what might be felt to be an irrational constellation, namely a combination of long pauses, an indicator of processes of selection and planning, and long " phrases ", an indicator of automatic action. If we also note the relative paucity of very short pauses (less than 0·5 sec.) in interviews we may be impressed by an all-or-none quality which is not shared in speech uttered under rational conditions. It would seem that the very short hesitations are a sign of a deliberate and cautious approach to verbal action even where choice is easy or already made. This is in keeping with the high proportion of short pauses in the well-learned reproductions (D 7, S 7) and the decreasing length of pauses with each repetition of a text already selected (Goldman-Eisler, 1961a). It is reasonable to argue that short pauses cannot be sustained in states of emotional excitement once speech action has got under way. Speech divided by pauses of less than 0·5 sec. would be speech not so much interrupted as speech slowed down, and we know that nothing requires central control as much as the slowing down of activity in progress. The incompatibility of distributing movement over a longer span of time with states of emotional excitement has been demonstrated by Luria (1932).

With this in mind it might be useful to contemplate the function of the longer pauses which are characteristic of interviews. If we consider that the verbal sequences which succeed them are mainly emotionally charged reminiscences or impulsive expressions of emotional attitudes of a length matched only by well-learned speech products (Goldman-Eisler, 1961b) we might be exercised by the following problem: If we are dealing here with largely automatic speech, i.e. highly predictable speech sequences, familiar to the speaker, how are we to explain the length of pauses, if pauses are indicators of selection and verbal planning?

I have previously referred to three types of decision which might be responsible for the delay of speech action (Goldman-Eisler, 1961). These were lexical decisions, decisions concerning structure and content decisions. I suggest that the first two, that is decisions strictly concerned with verbal planning, are of minor importance in psychiatric interviews. But that the interview situation does involve the interviewee in

serious decision-taking as far as content is concerned. And that the content decision is of a kind specific to such interviews, i.e. that it is not only a decision as to " what to say ", but also one of " whether " to verbalise contents which rise to the surface of consciousness in the course of the interview by virtue of its dynamics, as soon as the interviewee co-operates in the procedure. Unless guided and directed into a selective attitude towards his material (which was not the case in the interviews under discussion which were non-directive, the patient being left to his own associations) the material produced will be touched off by contiguous association.

The decision confronting the interviewee will therefore be whether to utter or to suppress contents presenting themselves more or less automatically. Much of the pausing in psychiatric interviews should account for conflict of this kind and the succession of a long pause by a long and fluent verbal sequence would indicate that the decision was one largely of whether to open the flood gates of surging material and to what extent to contain it rather than how to formulate it ; on the other hand, when long pauses are followed by statements structured by short pauses into a series of shorter phrases, we would suspect the decisions involved to be largely lexical and structural. Thus the combined measure of pause and phrase length is necessary to appreciate more specifically the processes underlying speech utterance.

References

Goldman-Eisler, F. (1958a). Speech production and the predictability of words in context. *Quart. J. exp. Psychol.*, **10**, 96.

Goldman-Eisler, F. (1958b). The predictability of words in context and the length of pauses in speech. *Language and Speech*, **1**, 226.

Goldman-Eisler, F. (1961a). Hesitation and information in speech. In *Information Theory : Proceedings of the 4th London Symposium on Information Theory* (London), 162.

Goldman-Eisler, F. (1961b). Continuity of speech utterance, its determinants and its significance. *Language and Speech*, **4**, 220.

Luria, A. R. (1932). The Nature of Human Conflict (New York).

INDEX TO VOLUME 4, 1961

AUTHORS

PAPERS

ADDRESSES OF CONTRIBUTORS TO VOL. 4, PART 4

BASTIAN, Jarvis, Center for Advanced Studies in the Behavioral Sciences, Stanford, California, U.S.A.

BLACK, Professor John W., Department of Speech, 154 N. Oval Drive, Ohio State University, Columbus 10, Ohio, U.S.A.

EIMAS, Peter, University of Connecticut, Storrs, Connecticut, U.S.A.

GOLDMAN-EISLER, Dr. Frieda, Department of Phonetics, University College London, Gower Street, London, W.C.1, England.

HARRIS, Dr. Katherine S., Haskins Laboratories, 305 East 43rd Street, New York 17, N.Y., U.S.A.

LIBERMAN, Dr. Alvin M., University of Connecticut, Storrs, Connecticut, U.S.A.

LISKER, Dr. Leigh, University of Pennsylvania, Philadelphia 4, Pennsylvania, U.S.A.

PETERSON, Dr. Gordon E., Communication Sciences Laboratory, University of Michigan, Ann Arbor, Michigan, U.S.A.

Printed by Clare o' Molesey Ltd. (T.U.), West Molesey, Surrey.